Genet. Res., Camb. (1996), **68**, *pp.* 249–257

Deviation from Hardy–Weinberg proportions in finite populations

JINLIANG WANG
College of Animal Science, Zhejiang Agricultural University, Hangzhou 310029, The People's Republic of China

(*Received 17 June 1996 and in revised form 27 August 1996*)

Summary

For a finite diploid population with no mutation, migration and selection, equations for the deviation of observed genotype frequencies from Hardy–Weinberg proportions are derived in this paper for monoecious species and for autosomal and sex-linked loci in dioecious species. It is shown that the genotype frequency deviation in finite random-mating populations results from the difference between the gene frequencies of male and female gametes, which is determined by two independent causes: the gene frequency difference between male and female parents and the sampling error due to the finite number of offspring. Previous studies have considered only one of the causes and the equations derived by previous authors are applicable only in the special case of random selection. The general equations derived here for both causes incorporate the variances and covariances of family size and thus they reduce to previous equations for random selection. Stochastic simulations are run to check the predictions from different formulae. Non-random mating and variation in census size are considered and the applications of the derived formulae are exemplified.

1. Introduction

In finite populations gene frequencies fluctuate randomly from generation to generation as a result of the sampling of a finite number of genes. Such populations will not attain Hardy–Weinberg equilibrium until one allele is fixed ($q = 1$) and all other alleles are completely lost (and thus genetic drift is absent). Before the equilibrium is reached, the observed genotype frequency deviates consistently from its expectation, calculated from the usual Hardy–Weinberg formula using the observed gene frequency. The relative deviation of the observed heterozygote frequency (H_o) from its expected value (H_e), $\alpha = (H_e - H_o)/H_e$ such that the observed heterozygote frequency is $H_o = 2q(1-q)(1-\alpha)$, in random-mating finite populations is the result of two independent random processes acting in each generation. The first is due to the finite number of progeny, which causes a difference in gene frequencies between male and female gametes that unite to form the progeny. The second is due to the finite number of parents, which causes a difference in gene frequencies between sexes in the parents. Thus, genetic drift in both parent and offspring generations will cause deviations of observed genotype frequencies from Hardy–Weinberg proportions.

The first process is considered by Kimura & Crow (1963) for a monoecious population with N individuals in each generation. The appropriate value of α is

$$\alpha = -\frac{1}{2N-1}, \tag{1}$$

which is extended to dioecious populations as

$$\alpha_s = -\frac{1}{2N_s-1}, \tag{2}$$

where N_s is the number of individuals of sex s and α_s is the deviation from Hardy–Weinberg proportions of sex s ($s = m$ or f). Many authors utilize (1) (Kimura & Crow, 1963; Caballero & Hill, 1992*a*, *b*; Crossa & Vencovsky, 1994; Wang, 1995) or (2) (Crow & Denniston, 1988; Caballero, 1994; Santiago & Caballero, 1995; Wang, 1996*a*, *b*) in various formulae to predict effective population size (Wright, 1938).

The second process is considered by Robertson (1965) for a population with N_m males and N_f females, and the prediction equation derived for α is

$$\alpha = -\frac{1}{8N_m} - \frac{1}{8N_f}. \tag{3}$$

Equation (3) is applied to the investigation of possible heterozygote superiority in plant and animal popu-

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

lations (Robertson, 1965) and also in predicting effective size (Caballero *et al.*, 1991).

In this paper, more appropriate equations to predict the relative genotype frequency deviation from Hardy–Weinberg proportions in finite populations of monoecious and dioecious species are obtained analytically and verified by stochastic simulations. For dioecious species, the deviation for each sex is considered for autosomal and sex-linked loci. It will be shown that both independent processes are involved in the genotype frequency deviation in dioecious species and thus should be considered simultaneously. And, for both processes, more general equations are derived which incorporate variances and covariances of family size and reduce to (1)–(3) for the special cases. In deriving the equations in this paper, we assume a random-mating population with constant size and structure, discrete generations and without mutation, migration and selection. Non-random mating and variation in population size are considered in the Discussion.

2. Theory

Genotype frequency deviation in finite random-mating populations results from the difference in the gene frequencies between two independent gamete samples: one of male gametes and the other of female gametes that unite to form the offspring. The difference will cause an apparent excess of heterozyotes in the progeny. The difference comes from two distinct sources, the first being the sampling error due to the finite size of male and female gamete sets, and the second being the difference in gene frequencies between male and female parents. It is clear that, for monoecious species, only the first source is responsible for the heterozygote excess.

Throughout this paper, the subscripts r and s will denote sex, m for male gametes or individuals and f for female gametes or individuals.

(i) *Monoecious species*

Consider a two-allele locus in a randomly fertilizing population of N individuals in each generation. If the gene frequencies of male and female gametes that unite to form offspring in generation t are q_m and q_f respectively, then the observed heterozygote frequency in generation t will be $H_o = q_m(1-q_f)+q_f(1-q_m)$, the mean gene frequency will be $q = \frac{1}{2}(q_m+q_f)$, and the expected heterozygote frequency from the Hardy–Weinberg formula will be $H_e = 2q(1-q) = q_m+q_f-\frac{1}{2}(q_m+q_f)^2$. There will then be an apparent excess of heterozygotes $H_e-H_o = -\frac{1}{2}(q_m-q_f)^2$. Thus any difference in gene frequencies between male and female gametes will cause an excess of heterozygotes, as first noted by Robertson (1965) for a dioecious population.

The difference between q_m and q_f depends on the sampling method and the heterozygosity of the parents in generation $t-1$. Let x_i be the gene frequency of parent i (thus, x_i is 0, $\frac{1}{2}$ or 1 if it carries zero, one or two copies of the allele, respectively), then the gene frequency in gametes of sex s is

$$q_s = \frac{1}{N}\sum_{i=1}^{N}\left(n_{si}x_i + \sum_{j=1}^{n_{si}}\delta_{ij}\right),$$

where n_{si} is the number of gametes of sex s produced by parent i, and δ_{ij} is the difference in gene frequency between the jth sampled gene and its parental value x_i, i.e. δ_{ij} is zero if the parent is a homozygote or $\pm\frac{1}{2}$ with equal probabilities if a heterozygote. The expected difference, $E(q_m-q_f)$, is zero. Thus the expected absolute genotype frequency deviation is obtained as

$$\begin{aligned}E(H_e-H_o) &= -\tfrac{1}{2}E(q_m-q_f)^2\\ &= -\tfrac{1}{2}\{E(q_m-q_f)^2-[E(q_m-q_f)]^2\}\\ &= -\tfrac{1}{2}V(q_m-q_f)\\ &= -\frac{1}{2N^2}\left\{V\left[\sum_{i=1}^{N}(n_{mi}x_i-n_{fi}x_i)\right]\right.\\ &\quad \left.+V\left[\sum_{i=1}^{N}\left(\sum_{j=1}^{n_{mi}}\delta_{ij}-\sum_{j=1}^{n_{fi}}\delta_{ij}\right)\right]\right\}.\end{aligned}$$

Since gene frequencies (x_i), numbers of gametes per parent (n_{si}) and Mendelian sampling terms (δ_{ij}) are uncorrelated, we have

$$\begin{aligned}E(H_e-H_o) = -\frac{1}{2N}\{&V(x_i)[V(n_{mi})-2\,\mathrm{Cov}(n_{mi},n_{fi})\\ &+V(n_{fi})]+2V(\delta_{ij})\}\end{aligned}$$

approximately, ignoring terms in $1/N^2$ relative to $1/N$ that are introduced by correlations among the n_{si} and x_i since their sums are fixed.

The variance of the gene frequency in the parents is $V(x_i) = q'(1-q')(1+\alpha')/2$, where $q' = 1/N\sum_{i=1}^{N}x_i$ is the gene frequency and α' is the deviation from Hardy–Weinberg proportions in generation $t-1$. The variance due to segregation, $V(\delta_{ij})$, is equal to the product of the frequency of heterozygotes, $2q'(1-q')(1-\alpha')$, and the variance generated from them, $\frac{1}{4}$, that is $V(\delta_{ij}) = q'(1-q')(1-\alpha')/2$. The variance of the total number of gametes per parent is $\sigma^2 = V(n_{mi}+n_{fi})$ and the covariance between the numbers of male and female gametes per parent is $\sigma_{mf} = \mathrm{Cov}(n_{mi},n_{fi})$. Substituting these relations into the expression for $E(H_e-H_o)$, we arrive at

$$\begin{aligned}E(H_e-H_o) = -\frac{q'(1-q')}{4N}&[(\sigma^2-4\sigma_{mf})\\ &\times(1+\alpha')+2-2\alpha'].\end{aligned}$$

The deviation from Hardy–Weinberg proportions resulting from genetic drift in the offspring generation is

$$\begin{aligned}\alpha &= \frac{E(H_e-H_o)}{H_e}\\ &= -\frac{q'(1-q')}{8Nq(1-q)}[(\sigma^2-4\sigma_{mf})(1+\alpha')+2-2\alpha'].\end{aligned}$$

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

In randomly fertilizing populations, α' is of order $1/N$ and

$$\frac{q'(1-q')}{q(1-q)} = 1 + o(1/N) \approx 1.$$

Thus neglecting terms in $1/N^2$ we get the approximate expression

$$\alpha = -\frac{2+\sigma^2-4\sigma_{mf}}{8N}. \tag{4}$$

From (4) we can see that not only the size of the offspring population, but also the variance and covariance of family size are important in determining the value of α. The larger the variance and the smaller the covariance, the greater is the relative genotype frequency deviation from Hardy–Weinberg proportions. If gametes are sampled randomly and independently from the parents, the distribution of the number of gametes per parent will be binomial. In such a case, $\sigma^2 = 2-2/N$ and $\sigma_{mf} = 0$. However, throughout this paper we neglect all second- and higher-order terms of $1/N$ and thus approximate the binomial distribution by the Poisson. Therefore, we have $\sigma^2 = 2$ and $\sigma_{mf} = 0$ approximately for the selection scheme. Inserting these values into (4) yields $\alpha = -1/(2N)$, which is the same as Kimura & Crow's (1963) equation (1) omitting $1/N^2$. For the special case of random selection, a comparison between (1) and (4) shows that (1) is a little more precise than (4), especially for very small populations. For example, the observed values of α from stochastic simulations are $-0{\cdot}145 \pm 0{\cdot}005$, $-0{\cdot}091 \pm 0{\cdot}003$ and $-0{\cdot}066 \pm 0{\cdot}003$ for $N = 4$, $N = 6$ and $N = 8$ respectively, while the expectations are $-0{\cdot}125$, $-0{\cdot}083$ and $-0{\cdot}063$ from (4) and $-0{\cdot}143$, $-0{\cdot}091$ and $-0{\cdot}067$ from (1). The difference between (1) and (4) for random selection is expected because second- and higher-order terms of $1/N$ are omitted for simplicity in deriving (4).

If each parent contributes exactly one male gamete and one female gamete to the next generation ($\sigma^2 = \sigma_{mf} = 0$), then the deviation is $\alpha = -1/(4N)$ from (4) – half the value for Poisson distribution of family size.

Wright (1938, 1939) derived the effective size of a randomly fertilizing population, which is

$$N_e = \frac{4N-2}{2+\sigma^2}. \tag{5}$$

Inserting the relation into (4) yields

$$\alpha = -\frac{1}{2N_e} + \frac{\sigma_{mf}}{2N} \tag{6}$$

approximately. Equation (6) shows that the relative excess of heterozygotes is equal to the rate of inbreeding ($\Delta F = 1/2N_e$) if the covariance between the numbers of male and female gametes per parent is zero. The deviation is related more closely to effective size than to census size. When $\sigma_{mf} \neq 0$, there is a difference between the (absolute) values of α and ΔF; and the larger the departure of the covariance from zero, the greater is the difference. This is because $|\alpha|$ measures the variance in the difference between the gene frequency of male gametes and that of female gametes, while ΔF measures the variance of change in the average gene frequency of male and female gametes between generations. If, for example, the numbers of male and female gametes contributed per parent are positively correlated ($\sigma_{mf} > 0$), gametes of separate sexes are more likely to come from the same parent and thus the gene frequency difference between male and female gametes will be decreased, while more gametes irrespective of sex are more likely to come from fewer parents and thus the average gene frequency change between generations will be increased. Therefore, $|\alpha|$ is decreased and ΔF increased for $\sigma_{mf} > 0$ compared with $\sigma_{mf} = 0$.

(ii) *Dioecious species*

We consider only autosomal loci in this section. Sex-linked loci are dealt with in the next one. We assume that the population consists of N_m males and N_f females in each generation, and that each male mates at random with an equal number, $R = N_f/N_m$ (R being an integer), of females.

If the gene frequencies in male and female gametes that come from generation $t-1$ and unite to form female individuals in generation t are $q_{mf,t-1}$ and $q_{ff,t-1}$ respectively, then the observed and expected heterozygote frequencies of females in generation t are

$$H_{of,t} = q_{mf,t-1}(1-q_{ff,t-1}) + q_{ff,t-1}(1-q_{mf,t-1}), \tag{7a}$$

$$H_{ef,t} = 2q_{f,t}(1-q_{f,t}), \tag{7b}$$

where $q_{f,t} = \frac{1}{2}(q_{mf,t-1} + q_{ff,t-1})$ is the mean gene frequency of female offspring. Let $D_{t-1} = q_{mf,t-1} - q_{ff,t-1}$, then we have $q_{mf,t-1} = q_{f,t} + \frac{1}{2}D_{t-1}$ and $q_{ff,t-1} = q_{f,t} - \frac{1}{2}D_{t-1}$. Substituting these relations into (7*a*) yields

$$H_{of,t} = 2q_{f,t}(1-q_{f,t}) + \tfrac{1}{2}D^2_{t-1}. \tag{7c}$$

The expectation of D_{t-1} is $E(D_{t-1}) = q_{m,t-1} - q_{f,t-1}$, where $q_{s,t-1}$ is the gene frequency of individuals of sex s in generation $t-1$, and the expectation of D^2_{t-1} is

$$E(D^2_{t-1}) = V(D_{t-1}) + (q_{m,t-1} - q_{f,t-1})^2. \tag{7d}$$

The variance of D_{t-1} is

$$V(D_{t-1}) = V(q_{mf,t-1} - q_{ff,t-1})$$

$$= V\left[\frac{1}{N_f}\sum_{i=1}^{N_m}\left(n_{mfi}x_{mi} + \sum_{j=1}^{n_{mfi}}\delta_{ij}\right) - \frac{1}{N_f}\sum_{i=1}^{N_f}\left(n_{ffi}x_{fi} + \sum_{j=1}^{n_{ffi}}\delta_{ij}\right)\right],$$

where n_{sfi} and x_{si} are the number of daughters and the gene frequency of individual i of sex s in generation $t-1$. Using a procedure similar to the monoecious

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

case and noting that all covariance terms are zero or of order $1/N_s^2$, we get

$$V(D_{t-1}) = \frac{q_{m,t-1}(1-q_{m,t-1})}{2N_f}\left[\left(\frac{N_m}{N_f}\right)\sigma_{mf}^2(1+\alpha'_m)+1-\alpha'_m\right] + \frac{q_{f,t-1}(1-q_{f,t-1})}{2N_f}[\sigma_{ff}^2(1+\alpha'_f)+1-\alpha'_f], \quad (7e)$$

where α'_s is the deviation from Hardy–Weinberg proportions in parents of sex s.

The expectation of the difference in gene frequency between male and female parents, $E(q_{m,t-1}-q_{f,t-1})$, should be equal to zero. Thus we have

$$E(q_{m,t-1}-q_{f,t-1})^2 = V(q_{m,t-1}-q_{f,t-1}) = V\left\{\frac{1}{N_m}\sum_{i=1}^{N_m}\sum_{j=1}^{R}\left[n_{fmij}\left(\frac{x'_{mi}+x'_{fij}}{2}\right)+\sum_{l=1}^{n_{fmij}}\left(\frac{\delta_{mijl}+\delta_{fijl}}{2}\right)\right] - \frac{1}{N_f}\sum_{i=1}^{N_m}\sum_{j=1}^{R}\left[n_{ffij}\left(\frac{x'_{mi}+x'_{fij}}{2}\right)+\sum_{l=1}^{n_{ffij}}\left(\frac{\delta_{mijl}+\delta_{fijl}}{2}\right)\right]\right\}, \quad (8)$$

where n_{fsij} is the number of individuals of sex s contributed by the jth female grandparent mated with the ith male grandparent, x'_{mi} and x'_{fij} are the gene frequencies of male i and its mate j ($j=1$ to R) in generation $t-2$, δ_{mijl} (δ_{fijl}) is the difference in gene frequency between the kth ($k=(j-1)R+l$ and $k=l$ for male and female grandparents respectively) sampled gene and its parental value $x'_{mi}(x'_{fij})$.

From (8) we can obtain, by a procedure similar to that given by Wang (1996*b*) and the derivation of (7*e*), the approximate expression

$$E(q_{m,t-2}-q_{f,t-2})^2 = \frac{q_{m,t-2}(1-q_{m,t-2})}{8N_m}\left\{\left[\sigma_{mm}^2-2\left(\frac{N_m}{N_f}\right)\sigma_{mm,mf}+\left(\frac{N_m}{N_f}\right)^2\sigma_{mf}^2\right](1+\alpha''_m)+\left(1+\frac{N_m}{N_f}\right)(1-\alpha''_m)\right\} + \frac{q_{f,t-2}(1-q_{f,t-2})}{8N_f}\left\{\left[\sigma_{ff}^2-2\left(\frac{N_f}{N_m}\right)\sigma_{fm,ff}+\left(\frac{N_f}{N_m}\right)^2\sigma_{fm}^2\right](1+\alpha''_f)+\left(1+\frac{N_f}{N_m}\right)(1+\alpha''_f)\right\}, \quad (9)$$

where α''_s is the deviation from Hardy–Weinberg proportions in grandparents of sex s in generation $t-2$, $\sigma_{sm,sf}$ is the covariance between the numbers of male and female individuals per grandparent of sex s, $q_{s,t-2}$ is gene frequency in grandparents of sex s.

From (7*b*)–(7*e*) and (9), we obtain the heterozygote frequency deviation in female offspring shown in (10)

$$\alpha_f = (H_{ef,t}-H_{of,t})/H_{ef,t} = -\frac{A_{m,t-1}}{8N_f}\left[\left(\frac{N_m}{N_f}\right)\sigma_{mf}^2(1+\alpha'_m)+1-\alpha'_m\right]-\frac{A_{f,t-1}}{8N_f}[\sigma_{ff}^2(1+\alpha'_f)+1-\alpha'_f] - \frac{A_{m,t-2}}{32N_m}\left\{\left[\sigma_{mm}^2-2\left(\frac{N_m}{N_f}\right)\sigma_{mm,mf}+\left(\frac{N_m}{N_f}\right)^2\sigma_{mf}^2\right](1+\alpha''_m)+\left(1+\frac{N_m}{N_f}\right)(1-\alpha''_m)\right\} - \frac{A_{f,t-2}}{32N_f}\left\{\left[\sigma_{ff}^2-2\left(\frac{N_f}{N_m}\right)\sigma_{fm,ff}+\left(\frac{N_f}{N_m}\right)^2\sigma_{fm}^2\right](1+\alpha''_f)+\left(1+\frac{N_f}{N_m}\right)(1-\alpha''_f)\right\}, \quad (10)$$

where

$$A_{s,n} = \frac{q_{s,n}(1-q_{s,n})}{q_{f,t}(1-q_{f,t})} \quad (s=m \text{ or } f \text{ and } n=t-2 \text{ or } t-1).$$

Since in equilibrium populations $1-A_{s,n}$, α'_s and α''_s are of order $1/N_s$, an approximate expression can be derived from (10), neglecting second- and higher-order terms of $1/N_s$, as

$$\alpha_f = \alpha_p + \alpha_{fo}, \quad (11)$$

where α_p is the deviation caused by the difference in gene frequency between male and female parents,

$$\alpha_p = -\frac{1}{32N_m}\left[2+\sigma_{mm}^2-2\left(\frac{N_m}{N_f}\right)\sigma_{mm,mf}+\left(\frac{N_m}{N_f}\right)^2\sigma_{mf}^2\right] - \frac{1}{32N_f}\left[2+\sigma_{ff}^2-2\left(\frac{N_f}{N_m}\right)\sigma_{fm,ff}+\left(\frac{N_f}{N_m}\right)^2\sigma_{fm}^2\right] \quad (12)$$

and α_{fo} is the deviation caused by sampling the offspring,

$$\alpha_{fo} = -\frac{2+\sigma_{ff}^2+(N_m/N_f)\sigma_{mf}^2}{8N_f}. \quad (13)$$

Similarly, we can derive the deviation in male offspring. The general equation for offspring of sex s is

$$\alpha_s = \alpha_p + \alpha_{so}, \quad (14)$$

where α_p is given by (12) and α_{so} is

$$\alpha_{so} = -\frac{2+\sigma_{fs}^2/\mu_{fs}+\sigma_{ms}^2/\mu_{ms}}{8N_s}, \quad (15)$$

where $\mu_{rs}=N_s/N_r$ is the average number of offspring of sex s per parent of sex r.

If we do not distinguish the sexes of the offspring, an equation for the deviation caused by sampling the offspring of both sexes, α_o, can also be derived following the same procedure shown above. The equation is

$$\alpha_o = -\frac{1}{2N}-\frac{N_m\sigma_m^2+N_f\sigma_f^2}{4N^2}, \quad (16)$$

where $N=N_m+N_f$ and $\sigma_s^2=\sigma_{sm}^2+2\sigma_{sm,sf}+\sigma_{sf}^2$. The

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

expression for α_p is given by (12), no matter whether the offspring are males, females or both.

Hill (1972, 1979) has obtained an equation for effective size of a randomly mating dioecious population, which is

$$\frac{1}{N_e} = \frac{1}{16N_m}\left[2+\sigma^2_{mm}+2\left(\frac{N_m}{N_f}\right)\sigma_{mm,mf}+\left(\frac{N_m}{N_f}\right)^2\sigma^2_{mf}\right] + \frac{1}{16N_f}\left[2+\sigma^2_{ff}+2\left(\frac{N_f}{N_m}\right)\sigma_{fm,ff}+\left(\frac{N_f}{N_m}\right)^2\sigma^2_{fm}\right]. \quad (17)$$

Substituting (17) into (12) we get

$$\alpha_p = -\frac{1}{2N_e}+\frac{\sigma_{fm,ff}}{8N_m}+\frac{\sigma_{mm,mf}}{8N_f}. \quad (18)$$

The average deviation caused by sampling male and female offspring is $\bar{\alpha}_o = (\alpha_{mo}+\alpha_{fo})/2$. Inserting (15) and (17) into the expression we obtain

$$\bar{\alpha}_o = -\frac{1}{N_e}+\frac{\sigma_{fm,ff}}{8N_m}+\frac{\sigma_{mm,mf}}{8N_f}. \quad (19)$$

From (18) and (19) we get the deviation from Hardy–Weinberg proportions averaged over male and female offspring,

$$\bar{\alpha} = -\frac{3}{2N_e}+\frac{\sigma_{fm,ff}}{4N_m}+\frac{\sigma_{mm,mf}}{4N_f}. \quad (20)$$

We can see from (18)–(20) that the relationships among deviations (α_p, $\bar{\alpha}_o$ and $\bar{\alpha}$), effective size and covariances ($\sigma_{fm,ff}$ and $\sigma_{mm,mf}$) are in essence the same as in the monoecious case. Comparing (16) and (19), we see that the values of $\bar{\alpha}_o$ and α_o are generally different. Only when $N_m = N_f$ and a Poisson distribution of family size is $\bar{\alpha}_o$ equal to α_o.

Equation (12) or (18) is the genotype frequency deviation caused by the difference in gene frequencies between male and female parents, which is different from (3) derived by Robertson (1965). He assumed random selection of both male and female individuals where the number of offspring of each sex follows a Poisson distribution, and he did not consider the variance and covariance of family size. More practically, however, male and female offspring are not necessarily selected at random from the population, especially in domestic animal populations. The general equation derived here incorporates the variances and covariances of both male and female offspring per family. For the special case of random selection, we have $\sigma^2_{rs} = N_s/N_r$ and $\sigma_{sm,sf} = 0$, and (12) or (18) reduces to (3) approximately, as expected. The smaller the variances and the larger the covariances of family size, the smaller the value of α_p predicted from (12) or (18) compared with that predicted from (3). For the selection scheme proposed by Gowe *et al.* (1959) to achieve minimal inbreeding and genetic drift in control populations, $\sigma^2_{fm} = (N_m/N_f)(1-N_m/N_f)$ and $\sigma^2_{mm} = \sigma^2_{ff} = \sigma^2_{mf} = \sigma_{mm,mf} = \sigma_{fm,ff} = 0$. Substituting these into (12) or (18) yields

$$\alpha_p = -\frac{3}{32N_m}-\frac{1}{32N_f}, \quad (21)$$

which (in absolute value) is always smaller than that given by (3).

Kimura & Crow (1963) have considered the genotype frequency deviation generated from sampling a finite number of offspring and obtained (1). Equation (2) is a direct extension of (1) for dioecious populations. From our (15) it can be seen that (2) is correct only for the special case of Poisson distribution of offspring. In such a case, (15) reduces to $\alpha_{so} = -1/(2N_s)$, similar to Kimura & Crow's results. More generally, however, α_{so} is also dependent on the variance of family size and the heterozygosity of parents as well as on the number of offspring. If there are differences in fertility or viability, for example, even if these are not inherited, the variance of family size will be larger than the Poisson expectation and thus the deviation will be greater than that predicted from (2). On the contrary, in control populations the variance of family size and thus α_{so} can be minimized. If the numbers of males and females are equal ($N/2$) in each generation and an equal number of offspring are selected from each family (minimal inbreeding), then all variances and covariances of family size are zero and (15) reduces to $\alpha_{mo} = \alpha_{fo} = 1/(2N)$, which is about half the value predicted from (2).

The homozygosity of parents also influences the value of α_{so}. We may take two extreme situations as examples. In the first, we assume that all parents are heterozygotes, and thus the deviation in parents is -1. Inserting this value into (10) yields $\alpha_{fo} = -1/(2N_f)$, independent of the variances of family size. This is intuitively correct. The importance of family size variance increases with the variation in parent genotypes. As an example at the other extreme, we assume that all parents are homozygotes, and thus their genotype frequency deviation is 1. From (10) we immediately obtain $\alpha_{fo} = -[(N_m/N_f)\sigma^2_{mf}+\sigma^2_{ff}]/(4N_f)$, which, only for Poisson distribution of family size, reduces to $\alpha_{fo} = -1/(2N_f)$ approximately. The two parent populations in the examples are far from equilibrium. Their genotype deviations due to gene frequency difference in parents (α_p), sampling of offspring (α_{so}) and both (α_s) will quickly approach asymptotically the equilibrium values given by (12), (15) and (14).

(iii) *Sex-linked loci*

Deviations from Hardy–Weinberg proportions for sex-linked loci or haplo-diploid species refers only to the homogametic sex. We assume that the population consists of N_m males and N_f females in each generation ($N_m \leqslant N_f$), and that each male mates at random with

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

an equal number ($R = N_f/N_m$ being an integer) of females.

We first assume that females are the homogametic sex. Using a procedure presented in the Appendix, we can obtain the genotypic deviation generated from gene frequency difference between male and female parents and from sampling a finite number of offspring as

$$\alpha_p = -\frac{1}{12N_m}\left[2+\left(\frac{N_m}{N_f}\right)^2\sigma_{mf}^2\right] - \frac{1}{24N_f}\left[1+\sigma_{ff}^2-4\left(\frac{N_f}{N_m}\right)\sigma_{fm,ff}+4\left(\frac{N_f}{N_m}\right)^2\sigma_{fm}^2\right], \quad (22)$$

$$\alpha_{fo} = -\frac{1+\sigma_{ff}^2+2(N_m/N_f)\,\sigma_{mf}^2}{8N_f}, \quad (23)$$

respectively. Combining α_p and α_{fo} by (14) gives the total genotypic deviation from Hardy–Weinberg proportions.

Pollak (1980, 1990) has derived an expression for effective size for sex-linked loci in a random-mating population, which, in the discrete generation case, is

$$\frac{1}{N_e} = \frac{1}{9N_m}\left[1+2\left(\frac{N_m}{N_f}\right)^2\sigma_{mf}^2\right] + \frac{1}{9N_f}\left[1+\sigma_{ff}^2+2\left(\frac{N_f}{N_m}\right)\sigma_{fm,ff}+\left(\frac{N_f}{N_m}\right)^2\sigma_{fm}^2\right]. \quad (24)$$

Using (14) and (22)–(24), we can rewrite the equation for α_f as

$$\alpha_f = -\frac{3}{2N_e}+\frac{\sigma_{fm,ff}}{2N_m}. \quad (25)$$

For random selection of offspring, (22) and (23) reduce to $\alpha_p = -1/(3N_m)-1/(6N_f)$ and $\alpha_{fo} = -1/(2N_f)$, respectively. In this case a comparison between sex-linked and autosomal loci reveals that the deviation caused by gene frequency difference between male and female parents (α_p) for sex-linked loci is always much larger than that for autosomal loci. Even with equal numbers of male and female individuals, male parents are evidently more important than female parents in determining α_p for sex-linked loci. The deviation caused by sampling a finite number of offspring (α_{fo}) for sex-linked loci is the same as that for autosomal loci.

For equal family size selection, (22) and (23) reduce to $\alpha_p = -1/(3N_m)+1/(8N_f)$ and $\alpha_{fo} = -1/(8N_f)$ respectively. In this case the absolute value of α_{fo} for sex-linked loci is smaller than that for autosomal loci. The total deviation for sex-linked loci is $\alpha_f = -1/(3N_m)$, irrespective of the number of females and is larger (in absolute value) than that for autosomal loci when $N_f > \frac{27}{23}N_m$.

Now we consider species where males are the homogametic sex, as in poultry. The same population structure and mating system also gives (22)–(25), substituting m for f and f for m, respectively. For the special case of a Poisson distribution of family size, the equations reduce to $\alpha_p = -1/(3N_f)-1/(6N_m)$, $\alpha_{mo} = -1/(2N_m)$ and $\alpha_m = -2/(3N_m)-1/(3N_f)$. For equal family size they reduce to $\alpha_p = -1/(8N_m)-1/(12N_f)$ and $\alpha_{mo} = -3/(8N_m)+1/(4N_f)$, and the total deviation from Hardy–Weinberg proportions is $\alpha_m = -1/(2N_m)+1/(6N_f)$. Thus, for a given number of males, the larger the female number, the greater the absolute value of α_m.

3. Simulations

Stochastic simulations have been carried out to check the equations of the present study that are in disagreement with those of the previous studies. The simulated population consists of N_m males and N_f females in each generation, each male mating at random with an equal number of $R = N_f/N_m$ (R being an integer) females to produce the next generation. We consider two selection schemes. The first is random selection (RS), where the numbers of male and female offspring per family follow a Poisson distribution. The second is minimal inbreeding or equal family size selection (ES), where one son is selected at random from each male parent and one daughter from each female parent (Gowe *et al.*, 1959).

Every simulation is run for 20 generations and 10000 replicates. The population in generation one is obtained by sampling N_m male and N_f female individuals at random from an infinite base population with gene frequency 0·5 in Hardy–Weinberg equilibrium. Values of α_s are obtained by calculating for each generation the relative deviation of observed heterozygote frequency from its expectation with the Hardy–Weinberg assumption in individuals of sex s. If the gene is lost or fixed in the population in any generation, the replicate is terminated and only values of α_s in previous generations of the replicates are used. The simulated values of α_s are averaged for all generations after generation three, when an asymptote has been reached, and over all replicates.

Table 1 shows the observed values of α_s and predicted values from (2), (3) and (14). Clearly, predicted results from (14) derived in this paper are in very close agreement with observed values for both autosomal and sex-linked cases. Both Kimura & Crow's (1963) and Robertson's (1965) equations underestimate the deviation from Hardy–Weinberg proportions. A combination of the two equations gives an approximate estimation of α_s when family size is Poisson-distributed, but not in general.

4. Discussion

In deriving the equations we have assumed that the population census size and structure are constant over generations. These equations can easily be extended to populations with variable sizes. In such a case, the parameters in expressions for α_{so} and α_p refer to parent and grandparent populations, respectively.

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

Table 1. *Observed and predicted deviation from Hardy–Weinberg proportions for populations with N_m males and N_f females, random selection (RS) or equal family size selection (ES) and autosomal or sex-linked loci*

Loci	Population	Observed α_m	Predicted from equation (2)	(3)	(14)	Observed α_f	Predicted from equation (2)	(3)	(14)
Autosomal	$N_m=4$, $N_f=16$								
	RS	$-0{\cdot}171\pm0{\cdot}006$	$-0{\cdot}143$	$-0{\cdot}039$	$-0{\cdot}164$	$-0{\cdot}068\pm0{\cdot}004$	$-0{\cdot}032$	$-0{\cdot}039$	$-0{\cdot}070$
	ES	$-0{\cdot}127\pm0{\cdot}003$	$-0{\cdot}143$	$-0{\cdot}039$	$-0{\cdot}111$	$-0{\cdot}043\pm0{\cdot}002$	$-0{\cdot}032$	$-0{\cdot}039$	$-0{\cdot}041$
	$N_m=8$, $N_f=40$								
	RS	$-0{\cdot}083\pm0{\cdot}004$	$-0{\cdot}067$	$-0{\cdot}019$	$-0{\cdot}081$	$-0{\cdot}031\pm0{\cdot}002$	$-0{\cdot}013$	$-0{\cdot}019$	$-0{\cdot}031$
	ES	$-0{\cdot}060\pm0{\cdot}003$	$-0{\cdot}067$	$-0{\cdot}019$	$-0{\cdot}056$	$-0{\cdot}019\pm0{\cdot}001$	$-0{\cdot}013$	$-0{\cdot}019$	$-0{\cdot}019$
Sex-linked	$N_m=4$, $N_f=16$								
	RS	$-0{\cdot}192\pm0{\cdot}007$	$-0{\cdot}143$	$-0{\cdot}039$	$-0{\cdot}188$	$-0{\cdot}121\pm0{\cdot}003$	$-0{\cdot}032$	$-0{\cdot}039$	$-0{\cdot}125$
	ES	$-0{\cdot}132\pm0{\cdot}006$	$-0{\cdot}143$	$-0{\cdot}039$	$-0{\cdot}115$	$-0{\cdot}086\pm0{\cdot}003$	$-0{\cdot}032$	$-0{\cdot}039$	$-0{\cdot}083$
	$N_m=8$, $N_f=40$								
	RS	$-0{\cdot}091\pm0{\cdot}004$	$-0{\cdot}067$	$-0{\cdot}019$	$-0{\cdot}092$	$-0{\cdot}059\pm0{\cdot}003$	$-0{\cdot}013$	$-0{\cdot}019$	$-0{\cdot}058$
	ES	$-0{\cdot}062\pm0{\cdot}004$	$-0{\cdot}067$	$-0{\cdot}019$	$-0{\cdot}058$	$-0{\cdot}041\pm0{\cdot}002$	$-0{\cdot}013$	$-0{\cdot}019$	$-0{\cdot}042$

For monoecious species, the deviation in generation t can be obtained, following the derivation of (4) but considering census size in each generation, as

$$\alpha = -\frac{1/\mu_{t-1}+(\sigma^2_{t-1}-4\sigma_{mf,t-1})/\mu^2_{t-1}}{2N_{t-1}}, \tag{26}$$

where N_{t-1} and $\mu_{t-1}=2N_t/N_{t-1}$ are the number of individuals and the mean number of gametes contributed per individual in generation $t-1$, and σ^2_{t-1} and $\sigma_{mf,t-1}$ are the variance and covariance of the number of gametes per individual in generation $t-1$. For autosomal loci in dioecious species, (12) and (15) can also be extended, for variable census size, to

$$\alpha_p = -\frac{1}{2N_e}+\frac{\sigma_{fm,ff,t-2}}{8N_{f,t-2}\mu_{fm,t-2}\mu_{ff,t-2}} + \frac{\sigma_{mm,mf,t-2}}{8N_{m,t-2}\mu_{mm,t-2}\mu_{mf,t-2}}. \tag{27}$$

where

$$\frac{1}{N_e} = \frac{1}{16N_{m,t-2}} \times\left[\frac{2}{\mu_{mm,t-2}}+\frac{\sigma^2_{mm,t-2}}{\mu^2_{mm,t-2}}+\frac{2\sigma_{mm,mf,t-2}}{\mu_{mm,t-2}\mu_{mf,t-2}}+\frac{\sigma^2_{mf,t-2}}{\mu^2_{mf,t-2}}\right] +\frac{1}{16N_{f,t-2}}\left[\frac{2}{\mu_{ff,t-2}}+\frac{\sigma^2_{ff,t-2}}{\mu^2_{ff,t-2}}+\frac{2\sigma_{fm,ff,t-2}}{\mu_{fm,t-2}\mu_{fm,t-2}}+\frac{\sigma^2_{fm,t-2}}{\mu^2_{fm,t-2}}\right] \tag{28}$$

(Wang, 1996*b*) and

$$\alpha_{so} = -\frac{1}{8N_{s,t}}\left[2+\frac{\sigma^2_{ms,t-1}}{\mu_{ms,t-1}}+\frac{\sigma^2_{fs,t-1}}{\mu_{fs,t-1}}\right], \tag{29}$$

respectively, where $N_{s,n}$ is the number of individuals of sex s in generation n, $\mu_{rs,n}=N_{s,n+1}/N_{r,n}$ and $\sigma^2_{rs,n}$ are the mean and variance of the number of offspring of sex s per parent of sex r in generation n, $\sigma_{rm,rf,n}$ is the covariance between the numbers of male and female offspring per parent of sex r in generation n. Similarly we can also get the equations for sex-linked loci incorporating variation in census size.

Another assumption made in the derivation is random mating. For non-random mating populations, an additional deviation from Hardy–Weinberg proportions results from inbreeding. For a monoecious population with partial self-fertilization proportion β, the deviation due to inbreeding or non-random mating, α_i, is $\alpha_i=\beta/(2-\beta)-(1+\sigma_{mf})/(2N-1-\sigma_{mf})$. When self-fertilization occurs in a random proportion, $\beta=(1+\sigma_{mf})/N$ and $\alpha_i=0$. In dioecious populations of equal numbers of male and female individuals ($N/2$) with a full-sib mating proportion β, the corresponding value of α_i is

$$\alpha_i = \beta/(4-3\beta)-(1+\sigma_{mf})/(2N-3-3\sigma_{mf}),$$

where σ_{mf} is the covariance between the numbers of male and female offspring per parent. Again for random mating we have $\beta=2(1+\sigma_{mf})/N$ and $\alpha_i=0$. For other systems of partial inbreeding, α_i can also be obtained analogously from the equations for equilibrium inbreeding coefficient derived by Hedrick & Cockerham (1986) for autosomal loci and by Wang (1996*a*) for sex-linked loci. The total deviation from Hardy–Weinberg proportions for non-random mating populations is $\alpha_s=\alpha_i+\alpha_p+\alpha_{so}$, approximately.

The equations derived here are useful in calculating gene frequencies from recessive homozygote frequencies in small populations. If, for example, in a dairy cattle population with four sires and 1000 dams each generation, a particular kind of recessive abnormality occurs at a proportion of 0·25%, then the recessive gene frequency is about 5% and the frequency of the recessive gene 'carriers' in normal individuals is 9·5% from Hardy–Weinberg law. However, because of the small number of sires, the population will have an apparent excess of heterozygotes. Assuming random mating and random selection of offspring, the deviation from Hardy–

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

Weinberg proportions will be $-0{\cdot}032$ approximately from (14). From the relation for recessive homozygote frequency $G = q^2 + q(1-q)\alpha$, we get

$$q = \frac{-\alpha + \sqrt{[\alpha^2 + 4(1-\alpha)G]}}{2(1-\alpha)} = 6{\cdot}7\,\%$$

and the frequency of 'carriers' is about 12·9 %, both being evidently larger than those expected from Hardy–Weinberg law. If the abnormality is determined by a recessive sex-liked gene, then its frequency and the 'carrier' frequency, obtained by using a deviation value of $-0{\cdot}084$ calculated from (22) and (23), are 10·1 and 18·3 % respectively, which are twice as large as the values from Hardy–Weinberg expectations.

The equations derived herein can also be used to estimate effective population size (N_e) from gene and genotype frequency data. Falconer (1981, pp. 68–69) listed the data from a mouse experiment. The mouse population consisted of 18 lines, all originating from the same random-bred base and all maintained by minimal inbreeding with eight pairs of parents mated in each of the 27 generations. The data consist of gene and genotype frequencies at five polymorphic enzyme loci in each of the lines. The inbreeding coefficient (F) at generation 27 is calculated from heterozygote frequency and the effective size is estimated from F. The exact effective size, calculated from the exact inbreeding coefficient using the pedigree records, is 28·6. The estimated effective size from the heterozygote frequency data is 32·7 if the deviation from Hardy–Weinberg proportions in each line is not accounted for, and the corresponding value is 20·3 if a deviation value calculated from $\alpha = -1/(2N)$, as used by Falconer (1981), is utilized. More correctly, however, the value of α should be $-3/(4N)$ from (12) and (16) derived in this paper. Using $\alpha = -3/(4N)$, we can obtain the estimate of effective size from the same heterozygote frequency data, which turns out to be 29·3 and is in close agreement with the exact value.

Appendix. Deviation from Hardy–Weinberg proportions for a sex-linked locus

We assume females are the homogametic sex. For a sex-linked locus with two alleles, we can obtain the absolute deviation from Hardy–Weinberg proportions in generation t,

$$H_{ef} - H_{of} = -\tfrac{1}{2}D_{t-1}^2, \qquad \text{(A 1)}$$

where $D_{t-1} = q_{mf,t-1} - q_{ff,t-1}$ is the difference in gene frequency between male and female gametes that come from parents in generation $t-1$ and unite to form female offspring in generation t. The expectation of D_{t-1} is

$$E(D_{t-1}) = E(q_{mf,t-1} - q_{ff,t-1}) = q_{m,t-1} - q_{f,t-1} = d_{t-1},$$

where $q_{s,t-1}$ is the gene frequency in parents of sex s,

$$q_{m,t-1} = \frac{1}{N_m}\sum_{i=1}^{N_m}\sum_{j=1}^{R}\left(n_{fmij}x'_{fij} + \sum_{l=1}^{n_{fmij}}\delta_{fijl}\right),$$

$$q_{f,t-1} = \frac{1}{N_f}\sum_{i=1}^{N_m}\sum_{j=1}^{R}\left[n_{ffij}\left(\frac{x'_{mi}+x'_{fij}}{2}\right) + \frac{1}{2}\sum_{l=1}^{n_{ffij}}\delta_{fijl}\right], \qquad \text{(A 2)}$$

where n_{fsij} and x'_{fij} are explained in (8) and x'_{mi}, the gene frequency of male grandparent i in the generation $t-2$, takes only two possible values, one or zero.

Unlike the autosomal case, the expectation of $d_{t-1} = q_{m,t-1} - q_{f,t-1}$ is not zero. Since n_{fsij} and x'_{fij} or x'_{mi} are uncorrelated, we can obtain from (A 2) that $E(d_{t-1}) = -\frac{1}{2}(q_{m,t-2} - q_{f,t-2}) = -\frac{1}{2}d_{t-2}$, where $q_{s,t-2}$ is the gene frequency of individuals of sex s in generation $t-2$. Thus the expectation of $V(d_{t-1})$ is $E[V(d_{t-1})] = E\{E(d_{t-1}^2) - [E(d_{t-1})]^2\} = E(d_{t-1}^2) - \frac{1}{4}E(d_{t-2}^2)$. In equilibrium $d_{t-2} = d_{t-1}$ so that $E(d_{t-1}^2) = \frac{4}{3}V(d_{t-1})$. Substituting (A 2) into the expression and noting that $V(x'_{mi}) = q_{m,t-2}(1-q_{m,t-2})$, we can obtain, using a procedure similar to the derivation of (9), that

$$\begin{aligned} E(q_{m,t-1} - q_{f,t-1})^2 &= \frac{q_{m,t-2}(1-q_{m,t-2})}{3N_m}\left[\left(\frac{N_m}{N_f}\right)^2\sigma_{mf}^2\right] \\ &\quad + \frac{2q_{f,t-2}(1-q_{f,t-2})}{3N_f} \\ &\quad \times\left\{\left[\tfrac{1}{4}\sigma_{ff}^2 - \left(\frac{N_f}{N_m}\right)\sigma_{fm,ff} + \left(\frac{N_f}{N_m}\right)^2\sigma_{fm}^2\right]\right. \\ &\quad \left.\times(1+\alpha''_f) + \left(\frac{1}{4} + \frac{N_f}{N_m}\right)(1-\alpha''_f)\right\}. \end{aligned} \qquad \text{(A 3)}$$

The expectation of D_{t-1}^2 is

$$\begin{aligned} E(D_{t-1}^2) &= V(D_{t-1}) + [E(D_{t-1})]^2 \\ &= V(q_{mf,t-1} - q_{ff,t-1}) + (q_{m,t-1} - q_{f,t-1})^2, \end{aligned} \qquad \text{(A 4)}$$

where the gene frequencies in male and female gametes that unite to form female offspring in generation t, $q_{mf,t-1}$ and $q_{ff,t-1}$, are

$$q_{mf,t-1} = \frac{1}{N_f}\sum_{i=1}^{N_m}(n_{mfi}x_{mi}),$$

$$q_{ff,t-1} = \frac{1}{N_f}\sum_{i=1}^{N_f}\left(n_{ffi}x_{fi} + \sum_{j=1}^{n_{ffi}}\delta_{ij}\right). \qquad \text{(A 5)}$$

Since $V(x_{mi}) = q_{m,t-1}(1-q_{m,t-1})$,

$V(x_{fi}) = q_{f,t-1}(1-q_{f,t-1})(1+\alpha'_f)/2$

and $V(\delta_{ij}) = q_{f,t-1}(1-q_{f,t-1})(1-\alpha'_f)/2$, we can obtain, from (A 5), that

$$\begin{aligned} V(q_{mf,t-1} - q_{ff,t-1}) &= \frac{q_{m,t-1}(1-q_{m,t-1})}{N_m}\left[\left(\frac{N_m}{N_f}\right)^2\sigma_{mf}^2\right] \\ &\quad + \frac{q_{f,t-1}(1-q_{f,t-1})}{2N_f} \\ &\quad \times[\sigma_{ff}^2(1+\alpha'_f) + 1 - \alpha'_f]. \end{aligned} \qquad \text{(A 6)}$$

Using (A 1), (A 3), (A 4) and (A 6) and neglecting second-order terms of $1/N_s$, we thus get (22) and (23).

https://doi.org/10.1017/S0016672300034224 Published online by Cambridge University Press

I thank Professor W. G. Hill and two anonymous referees for many constructive comments on drafts of this paper.

References

Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* **73**, 657–679.

Caballero, A. & Hill, W. G. (1992*a*). Effective size of nonrandom mating populations. *Genetics* **130**, 909–916.

Caballero, A. & Hill, W. G. (1992*b*). Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics* **131**, 493–507.

Caballero, A., Keightley, P. D. & Hill, W. G. (1991). Strategies for increasing fixation probabilities of recessive mutations. *Genetical Research* **58**, 129–138.

Crossa, J. & Vencovsky, R. (1994). Implications of the variance effective population size on the genetic conservation of monoecious species. *Theoretical and Applied Genetics* **89**, 936–942.

Crow, J. F. & Denniston, C. (1988). Inbreeding and variance effective population numbers. *Evolution* **42**, 482–495.

Falconer, D. S. (1981). *Introduction to Quantitative Genetics*, 2nd edn. New York: Longman.

Gowe, R. S., Robertson, A. & Latter, B. D. H. (1959). Environment and poultry breeding problems. 5. The design of poultry control strains. *Poultry Science* **38**, 462–471.

Hedrick, P. W. & Cockerham, C. C. (1986). Partial inbreeding: equilibrium heterozygosity and the heterozygosity paradox. *Evolution* **40**, 856–861.

Hill, W. G. (1972). Effective size of populations with overlapping generations. *Theoretical Population Biology* **3**, 278–289.

Hill, W. G. (1979). A note on effective population size with overlapping generations. *Genetics* **92**, 317–322.

Kimura, M. & Crow, J. F. (1963). The measurement of effective population number. *Evolution* **17**, 279–288.

Pollak, E. (1980). Effective population numbers and mean times to extinction in dioecious populations with overlapping generations. *Mathematical Biosciences* **52**, 1–25.

Pollak, E. (1990). The effective population size of an age-structured population with a sex-linked locus. *Mathematical Biosciences* **101**, 121–130.

Robertson, A. (1965). The interpretation of genotypic ratios in domestic animal populations. *Animal Production* **7**, 319–324.

Santiago, E. & Caballero, A. (1965). Effective size of populations under selection. *Genetics* **139**, 1013–1030.

Wang, J. (1995). Exact inbreeding coefficient and effective size of finite populations under partial sib mating. *Genetics* **140**, 357–363.

Wang, J. (1996*a*). Inbreeding coefficient and effective size for an X-linked locus in non-random mating populations. *Heredity* **76**, 569–577.

Wang, J. (1996*b*). Inbreeding and variance effective sizes for nonrandom mating populations. *Evolution* (in press).

Wright, S. (1938). Size of population and breeding structure in relation to evolution. *Science* **87**, 430–431.

Wright, S. (1939). Statistical genetics in relation to evolution. *Exposés de Biométrie et de Statisque Biologique*. Paris: Herman & Cie.