Effective Size and *F*-Statistics of Subdivided Populations. II. Dioecious Species

Jinliang Wang

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

Manuscript received January 22, 1996
Accepted for publication May 2, 1997

ABSTRACT

For a dioecious diploid population subdivided into an arbitrary number of subpopulations, we have derived recurrence equations for the inbreeding coefficient and coancestries between individuals within and among subpopulations and formulas for effective size and *F*-statistics. Stable population size and structure, discrete generations, autosomal inheritance, and the island migration model are assumed, and arbitrary distributions of the numbers of male and female progeny per family, different numbers and variable migration rates of males and females are incorporated in our derivation. Some published equations for effective size and *F*-statistics for a subdivided population are shown to be incorrect because several incorrect probabilities are used in the derivation. A more general equation for effective size is obtained by finding eigenvalue solutions to the recurrence equations for inbreeding coefficient and coancestry in this article, which reduces to the simple and familiar expressions derived by previous authors for the special case of a single unsubdivided population. Our general expressions for *F*-statistics also reduce to the classical results of WRIGHT's infinite island model and its extensions. It is shown that population structure is important in determining effective size and *F*-statistics and should be recognized and incorporated into programs for genetic conservation and evolution.

SINCE the introduction of the concept of effective population size by WRIGHT (1931), much work has been done on the prediction (CABALLERO 1994; and references therein) and estimation (WAPLES 1989; NUNNEY and ELAM 1994) of this parameter, which is central to evolutionary and quantitative genetics because it measures the rate of genetic drift and inbreeding. Most of the previous inquires on predicting effective size, however, are concerned with a single unsubdivided population. There is evidence that many organisms are arrayed into subpopulations (SELANDER 1970; CHESSER 1983; FOLTZ and HOOGLAND 1983) maintained by means of intrinsic factors, such as behavioral segregation, or extrinsic factors, such as geographic distance and habitat fragmentation. Domestic animal populations are normally subdivided into herds distributed in different farms with genetic exchanges (usually sires or semen) among them. The applications for animal breeding about effects of migration on inbreeding in subdivided populations have been discussed (ROUX 1995). Thus in recent years, formulas for effective size have been developed for subdivided populations. CHESSER (1991a,b) described the recurrence equations for inbreeding coefficient and coancestry and derived equations for effective size for a subdivided dioecious population, assuming that each female produces exactly one male and one female progeny. Subsequently, the model has been extended to independent Poisson distributions of male and female offspring per family (CHESSER *et al.* 1993) and multiple paternity (SUGG and CHESSER 1994). CHESSER and his coworkers noted that, their expressions for effective size and *F*-statistics when applied to a single unsubdivided population, are sometimes in variance with previous work (WRIGHT 1969; CABALLERO and HILL 1992). Also, they did not give eigenvalue solutions for the asymptotic effective size and thus their expressions presented did not preserve concordance with the classical models when the same assumptions are made. In this article, we extend the previous model to include an arbitrary distribution of family size as well as different numbers of male and female individuals, varying numbers of subpopulations and a variable rate of migration by each sex. We will derive expressions for asymptotic effective size, which reduce to the familiar equations for the special case of a single unsubdivided population as obtained by many authors (HILL 1979; NAGYLAKI 1995), by finding the eigenvalue solutions for recurrence equations for the inbreeding coefficient and coancestry. More general expressions for *F*-statistics will also be obtained, which reduce to the classical result of KIMURA and CROW (1963), ROBERTSON (1965), WRIGHT (1969), and PROUT (1981). We will show that, because several incorrect probabilities used in the derivation, CHESSER and coworker's expressions for effective size and *F*-statistics are different, but also incorrect.

ASSUMPTIONS AND PARAMETER DEFINITIONS

We consider a population subdivided into s subpopulations, each consisting of N_m males and N_f females with

---

Address for correspondence: Jinliang Wang, Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT, United Kingdom.
E-mail: jinliang.wang@ed.ac.uk

the total subpopulation size being $N = N_m + N_f$. We assume a stable census population size and structure in each discrete generation, autosomal inheritance involving genes which do not affect viability or reproductive ability so that natural selection is not operating to eliminate them.

As for the monoecious case, we assume the island model of migration. Each subpopulation receives a proportion of d_m males and a proportion of d_f unmated females drawn at random from the whole population. After migration, each subpopulation reproduces by panmixia.

The probabilities that two individuals of separate sexes, taken at random from the same subpopulation (q_{mf}) and different subpopulations (q'_{mf}) after migration, come from the same subpopulation before migration are

$$\begin{aligned} q_{mf} &= (1-d_m)(1-d_f) \\ &\quad + (1-d_m)d_f/s + (1-d_f)d_m/s + d_m d_f/s \\ &= 1 - (d_m + d_f - d_m d_f)(1 - 1/s) \qquad (1) \end{aligned}$$

and

$$\begin{aligned} q'_{mf} &= (1-d_m)d_f/s + (1-d_f)d_m/s + d_m d_f/s \\ &= (d_m + d_f - d_m d_f)/s, \qquad (2) \end{aligned}$$

respectively. Similar to the monoecious case (WANG 1997), we can derive the probabilities that two individuals of the same sex v ($v = m$ or f), taken at random and without replacement from the same subpopulation (q_{vv}) and different subpopulations (q'_{vv}) after migration, come from the same subpopulation before migration

$$\begin{aligned} q_{vv} &= (1-d_v)\left(1 - \frac{N_v d_v}{N_v - 1}\right) \\ &\quad + \frac{2N_v d_v(1-d_v)}{s(N_v-1)} + \frac{d_v(N_v d_v - 1)^2}{(N_v-1)(sN_v d_v - 1)}, \qquad (3) \end{aligned}$$

$$q'_{vv} = \frac{2d_v(1-d_v)}{s} + \frac{d_v^2(N_v d_v - 1)}{sN_v d_v - 1}. \qquad (4)$$

For relatively large values of N_v, they can be simplified to

$$q_{vv} = 1 - (2d_v - d_v^2)(1 - 1/s), \qquad (5)$$

$$q'_{vv} = (2d_v - d_v^2)/s \qquad (6)$$

approximately. Henceforth we use (5) and (6) instead of (3) and (4) for large values of N_v or for ease of presentation.

The parameters F and α as defined in the monoecious case are also used here. As to the coancestry between individuals within subpopulations (θ), three identity parameters, θ_{mm}, θ_{mf}, and θ_{ff} for the coancestry of two males, one male and one female, and two females taken at random from the same subpopulation before migration, respectively, can be distinguished for a dioecious population. Assuming that each female parent produced exactly one male and one female progeny, CHESSER (1991a,b) used only two of the three parameters in his model, θ_{mm} for coancestry of individuals of the same sex ($\theta_{mm} = \theta_{ff}$) and θ_{mf} for coancestry between male and female individuals. If the sex of the progeny produced by each female parent is assumed to be determined randomly (each sex produced with a probability of $1/2$), then $\theta_{mm} = \theta_{ff} = \theta_{mf}$ and there is no need to distinguish the coancestry of the same or separate sexes. CHESSER *et al.* (1993) and SUGG and CHESSER (1994) adopted the simplified model. More generally, however, θ_{mm}, θ_{mf}, and θ_{ff} should be distinguished; and in natural populations and especially in breeding populations of domestic animals, the distribution of the numbers of male and female progeny per parent may be quite different. Thus we use all the three parameters in our model. The expected coancestry between two individuals irrespective of sex, θ, can be found as

$$\begin{aligned} \theta &= \frac{N_m(N_m-1)}{N(N-1)}\theta_{mm} + \frac{2N_m N_f}{N(N-1)}\theta_{mf} \\ &\quad + \frac{N_f(N_f-1)}{N(N-1)}\theta_{ff}. \qquad (7) \end{aligned}$$

Corresponding to our general model, the probabilities that two random progeny within subpopulations before migration come from the same male or female parent should also be distinguished according to the sexes of the progeny. Using a procedure similar to WANG (1997), we can obtain the probabilities that two individuals of sex v ($p_{u,vv}$) and separate sexes ($p_{u,mf}$), taken at random from the same subpopulation before migration, come from the same parent of sex u

$$p_{u,vv} = \frac{1}{N_v - 1}\left[\left(\frac{N_u}{N_v}\right)\sigma_{uv}^2 + \frac{N_v}{N_u} - 1\right], \qquad (8)$$

and

$$p_{u,mf} = \left(\frac{N_u}{N_m N_f}\right)\sigma_{u,mf} + \frac{1}{N_u}, \qquad (9)$$

where subscripts u and v denote sex (u, $v = m$ or f), σ_{uv}^2 denotes the variance of the number of offspring of sex v from a parent of sex u, and $\sigma_{u,mf}$ represents the covariance of the numbers of male and female offspring from a parent of sex u. Equations (8) and (9) are also derived by NAGYLAKI (1995). The observed variance and covariance are $S_{uv}^2 = \sigma_{uv}^2 N_u/(N_u - 1)$ and $S_{u,mf} = \sigma_{u,mf}N_u/(N_u - 1)$, respectively.

CHESSER and coworkers (CHESSER 1991a,b; CHESSER *et al.* 1993; SUGG and CHESSER 1994) considered the variance in the number of females mated to a male (σ_b^2) and defined a corresponding parameter ϕ (the probability that random females within subpopulations mate with the same male) used in their derivations. By the inclusion of parameter σ_b^2 and ϕ, their models seem to be more general than our model herein, in fact this is not the case. In our model we, in concordance with most previous models, use the parameters σ^2_{mm}, σ^2_{mf}, and $\sigma_{m,mf}$, which include implicitly in themselves the variance in the number of females mated to a male (σ^2_b).

Since in our model three coancestry between individuals within subpopulations are distinguished according to the sexes of the individuals (θ_{mm}, θ_{mf}, and θ_{ff}), the parameters F_{IS} and F_{ST} can also be distinguished as

$$F_{IS(vw)} = \frac{F - \theta_{vw}}{1 - \theta_{vw}}, \quad F_{ST(vw)} = \frac{\theta_{vw} - \alpha}{1 - \alpha} \qquad (10)$$

correspondingly (v, $w = m$ or f). The average values of F_{IS} and F_{ST} can also be defined as

$$F_{IS} = \frac{F - \theta}{1 - \theta}, \quad F_{ST} = \frac{\theta - \alpha}{1 - \alpha}, \qquad (11)$$

where θ is given by (7).

RECURRENCE EQUATIONS FOR IDENTITIES BY DESCENT

The inbreeding coefficient, F, of an offspring is equal to the coancestry of its parents. Since a random pair of parents comes from the same subpopulation with probability q_{mf} and from different subpopulations with probability $1 - q_{mf}$ before migration, we have

$$F_t = q_{mf}\theta_{mf,t-1} + (1 - q_{mf})\alpha_{t-1}. \qquad (12)$$

The coancestry between two offspring of separate sexes, θ_{mf}, should equal to the identity (by descent) probability of the two genes, one taken at random from each of the two offspring. Three cases can be distinguished:

1. *Both genes descend from male parents with probability* $^1/_4$: In such a case, they come from the same parent with probability $p_{m,mf}$ and from different male parents with probability $1 - p_{m,mf}$. For the latter case, the parents are from the same subpopulation and different subpopulations before migration with probabilities q_{mm} and $1 - q_{mm}$, respectively. Thus, the two genes are identical by descent with probability
$$\tfrac{1}{4}\{p_{m,mf}(1 + F_{t-1})/2 + (1 - p_{m,mf}) \times [q_{mm}\theta_{mm,t-1} + (1 - q_{mm})\alpha_{t-1}]\}. \qquad (13)$$
2. *One gene comes from a male parent and the other gene from a female parent with probability* $^1/_2$: Thus they are identical by descent with probability
$$\tfrac{1}{2}[q_{mf}\theta_{mf,t-1} + (1 - q_{mf})\alpha_{t-1}]. \qquad (14)$$
3. *Both genes are from female parents with probability* $^1/_4$: Like the derivation of (13), we can derive the probability that two genes are identical by descent
$$\tfrac{1}{4}\{p_{f,mf}(1 + F_{t-1})/2 + (1 - p_{f,mf}) \times [q_{ff}\theta_{ff,t-1} + (1 - q_{ff})\alpha_{t-1}]\}. \qquad (15)$$

Collecting Expressions 13–15, we obtain the recurrence equation for the coancestry between two random offspring of separate sexes as

$$\begin{aligned}\theta_{mf,t} = {}& \tfrac{1}{8}(p_{m,mf} + p_{f,mf})(1 + F_{t-1}) \\ & + \tfrac{1}{4}(1 - p_{m,mf})q_{mm}\theta_{mm,t-1} + \tfrac{1}{2}q_{mf}\theta_{mf,t-1} \\ & + \tfrac{1}{4}(1 - p_{f,mf})q_{ff}\theta_{ff,t-1} + \tfrac{1}{4}[4 - q_{mm} - 2q_{mf} \\ & - q_{ff} - (1 - q_{mm})p_{m,mf} - (1 - q_{ff})p_{f,mf}]\alpha_{t-1}. \qquad (16)\end{aligned}$$

Similarly we can derive the recurrence equations for $\theta_{mm,t}$, $\theta_{ff,t}$, and α_t, which are

$$\begin{aligned}\theta_{mm,t} = {}& \tfrac{1}{8}(p_{m,mm} + p_{f,mm})(1 + F_{t-1}) \\ & + \tfrac{1}{4}(1 - p_{m,mm})q_{mm}\theta_{mm,t-1} + \tfrac{1}{2}q_{mf}\theta_{mf,t-1} \\ & + \tfrac{1}{4}(1 - p_{f,mm})q_{ff}\theta_{ff,t-1} + \tfrac{1}{4}[4 - q_{mm} - 2q_{mf} \\ & - q_{ff} - (1 - q_{mm})p_{m,mm} - (1 - q_{ff})p_{f,mm}]\alpha_{t-1}, \qquad (17)\end{aligned}$$

$$\begin{aligned}\theta_{ff,t} = {}& \tfrac{1}{8}(p_{m,ff} + p_{f,ff})(1 + F_{t-1}) \\ & + \tfrac{1}{4}(1 - p_{m,ff})q_{mm}\theta_{mm,t-1} + \tfrac{1}{2}q_{mf}\theta_{mf,t-1} \\ & + \tfrac{1}{4}(1 - p_{f,ff})q_{ff}\theta_{ff,t-1} + \tfrac{1}{4}[4 - q_{mm} - 2q_{mf} \\ & - q_{ff} - (1 - q_{mm})p_{m,ff} - (1 - q_{ff})p_{f,ff}]\alpha_{t-1}, \qquad (18)\end{aligned}$$

and

$$\begin{aligned}\alpha_t = {}& \tfrac{1}{4}q'_{mm}\theta_{mm,t-1} + \tfrac{1}{2}q'_{mf}\theta_{mf,t-1} + \tfrac{1}{4}q'_{ff}\theta_{ff,t-1} \\ & + \tfrac{1}{4}(4 - q'_{mm} - 2q'_{mf} - q'_{ff})\alpha_{t-1}, \qquad (19)\end{aligned}$$

respectively.

Expressions 12 and 16–19 fully describe the transitions of coancestry or gene identities by descent for a subdivided dioecious population over successive generations. A matrix form of presentation is

$$\mathbf{S}_t = \mathbf{T}\mathbf{S}_{t-1} + \mathbf{C}, \qquad (20)$$

where

$$\mathbf{S}_t = \begin{pmatrix} F_t \\ \theta_{mm,t} \\ \theta_{mf,t} \\ \theta_{ff,t} \\ \alpha_t \end{pmatrix}, \quad \mathbf{C} = \tfrac{1}{8}\begin{pmatrix} 0 \\ p_{mm} \\ p_{mf} \\ p_{ff} \\ 0 \end{pmatrix},$$

$$\mathbf{T} = \tfrac{1}{8}\begin{pmatrix} 0 & 0 & 8q_{mf} & 0 & 8 - 8q_{mf} \\ p_{mm} & 2b_{mm} & 4q_{mf} & 2c_{mm} & 2(Q - p_{mm} + q_{mm}p_{m,mm} + q_{ff}p_{f,mm}) \\ p_{mf} & 2b_{mf} & 4q_{mf} & 2c_{mf} & 2(Q - p_{mf} + q_{mm}p_{m,mf} + q_{ff}p_{f,mf}) \\ p_{ff} & 2b_{ff} & 4q_{mf} & 2c_{ff} & 2(Q - p_{ff} + q_{mm}p_{m,ff} + q_{ff}p_{f,ff}) \\ 0 & 2q'_{mm} & 4q'_{mf} & 2q'_{ff} & 2Q' \end{pmatrix}. \qquad (21)$$

In matrixes **C** and **T**, $p_{vw} = p_{m,vw} + p_{f,vw}$, denoting the probability that a random pair of offspring of sexes v and w, respectively, come from the same parent, $b_{vw} = q_{mm}(1 - p_{m,vw})$, $c_{vw} = q_{ff}(1 - p_{f,vw})$, $Q = 4 - q_{mm} - 2q_{mf} - q_{ff}$, and $Q' = 4 - q'_{mm} - 2q'_{mf} - q'_{ff}$, with subscripts v and w denoting sexes ($v, w = m$ or f).

Generally male and female migration rates have differential effects on inbreeding coefficient and coancestry. Only when $N_m = N_f$ and the numbers of male and female offspring per family are in the same and independent distributions do male and female migrations have equal effects. Figure 1 depicts the effects of male and female migrations on the inbreeding coefficient of a population subdivided into 20 subpopulations, each consisting of 5 males and 50 females, with other parameters being $\sigma^2_{mm} = \sigma^2_{mf} = \sigma_{m,mf} = \sigma_{f,mf} = 0$ and $\sigma^2_{fm} = \sigma^2_{ff} = 2$. Figure 1 shows that the magnitude of effect of male migration is different from that of female migration. This is due to the differences in male and female numbers and in the distributions of progeny per male and female parent. Similar to the monoecious case, male or female migration decreases the inbreeding coefficients in initial generations and increases the inbreeding coefficients in later generations. Thus the smaller the migration rate, the lower the eventual inbreeding coefficient. For example, in Figure 1 the line for $d_m = 0.1$ will cross lines for $d_m = 0.2$, 0.3, and 0.4 in generations 260, 231, and 214, respectively when $d_f = 0$, while the line for $d_f = 0.1$ will cross lines for $d_f = 0.2$, 0.3, and 0.4 in generations 261, 244, and 239, respectively, when $d_m = 0$.

Let us now consider some special cases of the recurrence equations and make comparisons with previous results.

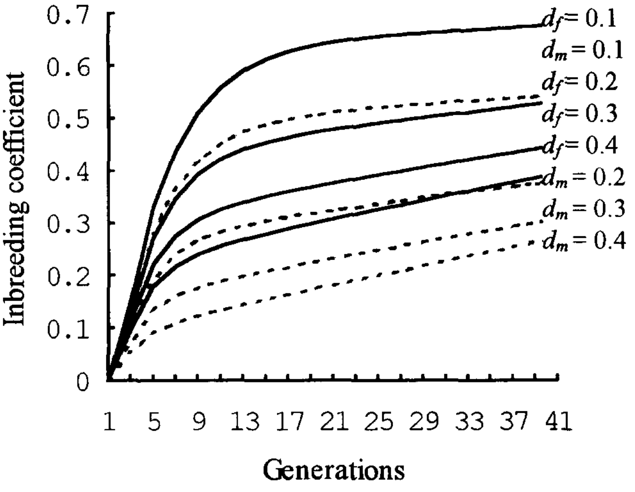

FIGURE 1.—The inbreeding coefficient over the first 41 generations for a population subdivided into 20 subpopulations, each consisting of 5 males and 50 females, with parameters $\sigma^2_{mm} = \sigma^2_{mf} = \sigma_{m,mf} = \sigma_{f,mf} = 0$ and $\sigma^2_{fm} = \sigma^2_{ff} = 2$. -----, effect of d_m with a constant value of d_f (0); ——, effect of d_f with d_m being constant (0).

Each female parent contributes exactly one son and one daughter to the next generation. Females do not migrate ($d_f = 0$) and males migrate randomly among the subpopulations ($d_m = 1$): The selection scheme implicitly confines that the numbers of males and females within subpopulations are equal ($N_m = N_f = n$). In such a case, we have $\sigma^2_{uv} = \sigma_{u,mf} = 0$, $p_{u,vv} = 0$, $p_{u,mf} = 1/n$, $q_{mf} = q'_{mf} = 1/s$, $q_{ff} = 1$, $q'_{ff} = 0$, and $q'_{mm} = q_{mm} = (n - 1)/(sn - 1)$ (from the exact Equation 3). With these stipulations (21) reduces to

$$\mathbf{S}_t = \begin{pmatrix} F_t \\ \theta_{mm,t} \\ \theta_{mf,t} \\ \alpha_t \end{pmatrix}, \quad \mathbf{C} = {}^1\!/_4 \begin{pmatrix} 0 \\ 0 \\ 1/n \\ 0 \end{pmatrix},$$

$$\mathbf{T} = \begin{pmatrix} 0 & 0 & y & 1 - y \\ 0 & \dfrac{1 + x}{4} & \dfrac{y}{2} & \dfrac{3 - x - 2y}{4} \\ \dfrac{1}{4n} & \dfrac{(n - 1)(1 + x)}{4n} & \dfrac{y}{2} & \dfrac{n(3 - x - 2y) - 1 + x}{4n} \\ 0 & \dfrac{x}{4} & \dfrac{y}{2} & \dfrac{4 - x - 2y}{4} \end{pmatrix}, \qquad (22)$$

where $x = (n - 1)/(sn - 1)$ and $y = 1/s$. Because of the special selection scheme, the coancestry between two males is always equal to the coancestry between two females ($\theta_{mm} = \theta_{ff}$), and thus they are all denoted as θ_{mm} and the variables in the matrix equations reduce to four. Equation (22) is also derived by CHESSER (1991a, Equations 19–21) for the case of equal numbers of males and females. Although his equations consider unequal numbers of different sexes, they are correct only when $N_m = N_f$ because of the restriction of the selection scheme and the incorrect definition of the probability (x) that two males taken at random are from the same subpopulation, $x = (N_f - 1)/(sN_f - 1)$ (CHESSER 1991a, Equation 8), which should be $x = (N_m - 1)/(sN_m - 1)$.

The selection scheme is the same as the above but both males and females are allowed to disperse with variable migration rates: In such a case, it can be shown that our transition matrix in (21) reduces to

$$\mathbf{T} = \begin{pmatrix} 0 & 0 & q_{mf} & 1 - q_{mf} \\ 0 & \dfrac{q_{mm} + q_{ff}}{4} & \dfrac{q_{mf}}{2} & \dfrac{Q}{4} \\ \dfrac{1}{4n} & \dfrac{(n-1)(q_{mm} + q_{ff})}{4n} & \dfrac{q_{mf}}{2} & \dfrac{Qn - 2 + q_{mm} + q_{ff}}{4n} \\ 0 & \dfrac{q'_{mm} + q'_{ff}}{4} & \dfrac{q'_{mf}}{2} & \dfrac{Q'}{4} \end{pmatrix}, \tag{23}$$

which is in variance with the transition matrix derived by CHESSER (1991b, Equation 5). In the derivation of his equations, the probabilities that two individuals of sex v taken at random from the same subpopulation and from different subpopulations after migration come from the same subpopulation before migration are mistakenly used as (CHESSER 1991b, in Equations A3, A5, A8, and A9)

$$q_{vv} = 1 - \left(1 - \frac{N_f - 1}{sN_f - 1}\right) d_v \tag{24}$$

and

$$q'_{vv} = \left(\frac{N_f - 1}{sN_f - 1}\right) d_v \tag{25}$$

respectively, while the correct expressions for q_{vv} and q'_{vv} should be given by (5) and (6), or more exactly by (3) and (4) shown in this article. Only when d_v equals zero or one, do (24) and (25) give correct values for q_{vv} and q'_{vv}; Otherwise, (24) always overestimates the value of q_{vv} and (25) underestimates the value of q'_{vv}. With a small number of subpopulations and intermediate values of migration rates, the errors made by (24) and (25) are great. For example, with $s = 2$, $d_v = 0.5$, and large N_f, the values of q'_{vv} are 0.25 and 0.375 approximately from (25) and (6), respectively.

The numbers of male and female offspring are in independent Poisson distributions and males and females disperse with migration rates d_m and d_f, respectively: In this case, $S^2_{uv} = N_v / N_u$ and $S_{u,mf} = 0$, and from (8) and (9) we find that $p_{u,vv} = p_{u,mf} = 1/N_u$. Thus it is clear that $\theta_{mm} = \theta_{mf} = \theta_{ff} = \theta$ and there is no need to distinguish the coancestry between individuals according to their sexes. With these stipulations (21) reduces to

$$\mathbf{S}_t = \begin{pmatrix} F_t \\ \theta_t \\ \alpha_t \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 0 \\ \dfrac{1}{8N_m} + \dfrac{1}{8N_f} \\ 0 \end{pmatrix},$$

$$\mathbf{T} = \begin{pmatrix} 0 & q_{mf} & 1 - q_{mf} \\ \dfrac{N}{8N_mN_f} & 1 - \dfrac{Q}{4} - \dfrac{q_{mm}}{4N_m} - \dfrac{q_{ff}}{4N_f} & \dfrac{Q}{4} - \dfrac{1 - q_{mm}}{4N_m} - \dfrac{1 - q_{ff}}{4N_f} \\ 0 & 1 - \tfrac{1}{4}Q' & \tfrac{1}{4}Q' \end{pmatrix}. \tag{26}$$

T and **C** in (26) are different from the corresponding expressions derived by CHESSER *et al.* [1993, Equations 12 and 13 where parameters ϕ, k, and σ^2_k take values $\phi = N_f / [N_m(N_f - 1)]$ and $k = \sigma^2_k = 1 + N_m/N_f$ for the special case]. In their derivations of the recurrence equations, the incorrect probabilities (24) and (25) are used and thus different and also incorrect recurrence equations result. And it is meaningless to include the variances and covariances of family size in their expressions, because the premise made by them ($\theta_{mm} = \theta_{mf} = \theta_{ff}$) implicitly means that the numbers of males and females must be in independent Poisson distributions. Following the methodology of CHESSER *et al.* (1993), SUGG and CHESSER (1994) extended their equations to include multiple paternity. The equations derived by them, however, are also incorrect because (24) and (25) are used.

EFFECTIVE SIZE

The matrix equation is exact and is easily programmed to provide numerical results for inbreeding coefficient and coancestry. The instantaneous rates of changes of the five gene identity parameters in $\mathbf{S}_t$ (ΔF_t, $\Delta\alpha_t$, $\Delta\theta_{mm,t}$, $\Delta\theta_{mf,t}$, and $\Delta\theta_{ff,t}$) and $\theta(\Delta\theta_t)$ and thus the six corresponding instantaneous effective sizes ($N_{eI,t}$, $N_{eV,t}$, $N_{e\theta(mm),t}$, $N_{e\theta(mf),t}$, $N_{e\theta(ff),t}$, and $N_{e\theta,t}$) can also be predicted using the matrix equation in a way similar to the monoecious case (WANG 1997). It is found, however, that the differences among the last four effective sizes are generally negligible except for the first two or three generations, thus only $N_{e\theta,t}$ is considered in the following numerical example.

For a population subdivided into 20 subpopulations, each consisting of 5 males and 50 females with male

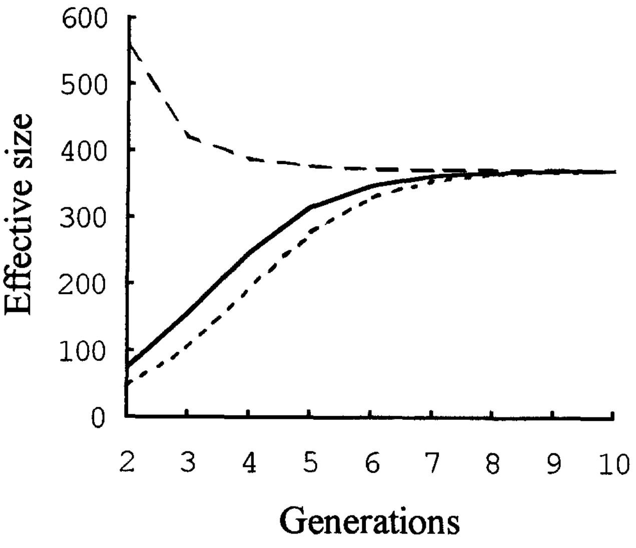

FIGURE 2.—Changes in the three effective sizes (N_{eI}, $N_{e\theta}$, and N_{eV} denoted by solid, dotted and dashed lines respectively) over successive generations for a population with parameters $s = 20$, $N_m = 5$, $N_f = 50$, $d_m = 0.8$, $d_f = 0$, $S^2_{mm} = S^2_{ff} = 1$, $S^2_{mf} = 1/S^2_{fm} = 10$, and $S_{m,mf} = S_{f,mf} = 0$.

and female migration rates being 0.8 and 0, respectively, and $S^2_{mm} = S^2_{ff} = 1$, $S^2_{mf} = 1/S^2_{fm} = N_f/N_m$, and $S_{u,mf} = 0$ (following Poisson distribution of family size), the changes in the effective sizes over the first 10 generations are shown in Figure 2. It is clear that, although the three effective sizes are quite different in initial generations, they will converge to the same value eventually. Results for other populations with different values of s, N_u, d_u, S^2_{uv}, and $S_{u,mf}$ are similar to those shown. Although the number of generations required to attain the asymptotic value of effective size may be different for various populations, the population must reach the steady state in the end so long as they are not completely subdivided ($d_m + d_f > 0$).

Numerical results for the asymptotic effective size can be obtained using (20). Analytical expressions for the equilibrium effective size can be derived by finding the dominant characteristic root of transition matrix **T**. The characteristic equation of matrix **T** may be reduced to

$$\begin{vmatrix} -4\lambda & 0 & q_{mf} & 0 & 1-2\lambda \\ p_{mm}+4\lambda & b_{mm}-4\lambda & 0 & c_{mm} & 1 \\ p_{mf}+4\lambda & b_{mf} & -2\lambda & c_{mf} & 1 \\ p_{ff}+4\lambda & b_{ff} & 0 & c_{ff}-4\lambda & 1 \\ 0 & q'_{mm} & q'_{mf} & q'_{ff} & 2-2\lambda \end{vmatrix} = 0. \quad (27)$$

Extending (27), we obtain the multinomial equation, from which the dominant root λ (the largest root lying between values zero and one) can be obtained and the value of effective size is calculated as $N_e = 0.5/(1-\lambda)$ (WANG 1997). From (8) and (9), we can see that $p_{u,vw}$ is of order $1/N_m$ or $1/N_f$. For relatively large subpopulation size (N_m, $N_f \gg 1$), second and higher orders of $p_{u,vw}$ can be omitted and (5) and (6) can be used in an approximate solution for λ. Following a similar procedure as for the monoecious case (WANG 1997), the effective size can be obtained as

$$N_e = \frac{16s}{p_{mm} + 2p_{mf} + p_{ff}} + \frac{2s(A_m + A_f)}{(d_m + d_f)(4 - d_m - d_f)(p_{mm} + 2p_{mf} + p_{ff})}, \quad (28)$$

where

$$\begin{aligned} A_m = {} & [(1 + 9d_f + 5d_m - 5d_m d_f - 2d_f^2)\,q_{mm} \\ & - (1 - d_m)(1 - 2d_m + d_f)/s]\,p_{m,mm} \\ & + [(1 - d_m)(1 - d_f)(5q_{mm} + q_{ff} - 4) \\ & - (2 + 2d_m + 2d_f - d_m^2 - d_f^2)/s]\,p_{m,mf} \\ & + (1 - d_f)[2(1 - d_f)\,q_{mm} - (1 - d_m)\,q_{ff} \\ & - (1 - 2d_f + d_m)/s]\,p_{m,ff} \quad (29) \end{aligned}$$

or

$$\begin{aligned} A_m \approx {} & [1 + 9d_f + 3d_m - (1 + d_f - 5d_m)/s]\,p_{m,mm} \\ & + 2[1 - 6d_m - 2d_f - (1 - 4d_m)/s]\,p_{m,mf} \\ & + [1 - 3d_m - d_f - (1 - 3d_m)/s]\,p_{m,ff} \quad (30) \end{aligned}$$

approximately, when d_m and d_f are sufficiently small for squares and products of these quantities to be ignored. A_f can also be expressed by (29) or (30) only to replace subscript m with f and f with m ($p_{u,fm} = p_{u,mf}$ because $\sigma_{u,fm} = \sigma_{u,mf}$ in Equation 9).

Expression 28 provides exact fits for the values of effective size obtained via iterations of Equation 20, as would be expected. For a single unsubdivided population ($s = 1$, $d_m = d_f = 1$), we get $q_{vw} = q'_{vw} = 1$ (from Equations 1–4). Substituting these and $p_{vw} = p_{m,vw} + p_{f,vw}$ into (28) yields

$$N_e = \frac{16}{p_{m,mm} + 2p_{m,mf} + p_{m,ff} + p_{f,mm} + 2p_{f,mf} + p_{f,ff}} \quad (31)$$

approximately, which was also derived by NAGYLAKI (1995) for the special case by a different method. When (8) and (9) are inserted into (31), we are led to HILL's (1979) formula for the variance effective population size as found by NAGYLAKI (1995). Thus our expression for effective size is equal to previous results by many authors when applied to a single isolated population.

The general expressions for effective size, (28) and (29), can be simplified considerably for the special case considered by CHESSER *et al.* (1993). Since the numbers of male and female progeny per family are in Poisson distributions, we obtain $p_{u,mm} = p_{u,mf} = p_{u,ff} = 1/N_u$ (where $u = m$ or f). Inserting this into (28) and (29) yields

$$N_e = \frac{4sN_m N_f}{N_m + N_f} + \frac{\begin{array}{l} s(1 - d_m)(1 - 2d_m + d_f)N_f \\ \quad + s(1 - d_f)(1 - 2d_f + d_m)N_m - N_m - N_f \end{array}}{\tfrac{1}{2}(N_m + N_f)(d_m + d_f)(4 - d_m - d_f)}. \quad (32)$$

Equations 28 and 32 give satisfactory predictions of effective size when both N_m and N_f are relatively large. If, however, either N_m or N_f is very small, they may result in a little bias because the approximate expressions (5) and (6) are used. A general equation for N_e using exact expressions (3) and (4) can also be derived from (27), it is, however, quite complicated and not given here. For the special case of Poisson distribution of family size, we can get the equation for N_e from transition matrix **T** in (26) as

$$N_e = \frac{4N_m N_f}{N_m + N_f} + \frac{\begin{array}{l} 4(4 - q_{mm} - 2q_{mf} - q_{ff})N_m N_f \\ \quad + 2(2q_{mm} - q_{mf})N_f + 2(2q_{ff} - q_{mf})N_m \end{array}}{(q'_{mm} + 2q'_{mf} + q'_{ff})(N_m + N_f)}. \quad (33)$$

For relatively large values of N_m and N_f, inserting (5) and (6) into (33) we also get (32) approximately, as expected.

A close inspection of (28) informs us that, for a given population, the smaller the migration rates ($d_m + d_f > 0$), the larger the effective size. The magnitude of the effect of migration rates on effective size depends mainly on the number of subpopulations and is independent of the census size of the subpopulation. The effects of s, d_m, and d_f on effective size are shown in Figure 3 for a population with fixed total size (2000), sex ratio ($N_f/N_m = 5$), Poisson distribution of family size ($S^2_{mm} = S^2_{ff} = 1$, $S_{m,mf} = S_{f,mf} = 0$, and $S^2_{mf} = 1/S^2_{fm} = 5$), and variable values of s, d_m, and d_f (we set $d_m = d_f$). It is clear that the largest value of N_e is achieved with low migration rates and a large number of subpopulations. If either d_m or d_f is large or s is small, the population behaves as a random mating unsubdivided one and population structure has little effect on effective size.

Male and female migration rates may have differential effects on effective size. Depending on the sex ratio and the variances and covariances of family size, the relative effect of d_m on effective size may be larger or smaller than that of d_f. When $N_m = N_f$ and the numbers of male and female offspring per family follow the same and independent distributions, d_m and d_f have equal effects. The influence of male and female migration rates on effective size is presented in Figure 4 for a population subdivided into 20 subpopulations, each of 10 males and 50 females, with independent Poisson distributions of male and female progeny per family ($S^2_{uv} = N_v/N_u$, $S_{u,mf} = 0$). Figure 4 shows that d_m has a larger effect on N_e than d_f in this case.

The expressions for effective size derived above are valid when migration exists. If the subpopulations are completely isolated ($d_m = d_f = 0$), $N_{eI,t}$, $N_{e\theta(mm),t}$, $N_{e\theta(mf),t}$, $N_{e\theta(ff),t}$, and $N_{e\theta,t}$ will converge to the same asymptotic value, the inbreeding effective size, while $N_{eV,t}$ tends to another equilibrium value, variance effective size. In such a case, the rate of inbreeding of the total population is the same as any one of its subpopulations. The inbreeding effective size of a subpopulation is also that of the total population. In this situation the matrixes (21) are reduced considerably for $\alpha = 0$, and following a similar procedure, we can derive the asymptotic inbreeding effective size, which is expressed by (31) approximately.

F-STATISTICS

Population subdivision and migration change the *F*-statistics considerably. The equations for the instantaneous *F*-statistics can be obtained from (10–12) and (16–19),

$$\frac{1}{2N_{eI,t}} = \frac{q_{mf}F_{ST(mf),t-1} - F_{IT,t-1}}{1 - F_{IT,t-1}}, \quad (34)$$

$$\begin{aligned} \frac{1}{2N_{e\theta(vw),t}} = \frac{1}{8(1 - F_{ST(vw),t-1})} [&p_{vw}(1 + F_{IT,t-1}) \\ &+ 2q_{mm}(1 - p_{m,vw})F_{ST(mm),t-1} + 4q_{mf}F_{ST(mf),t-1} \\ &+ 2q_{ff}(1 - p_{f,vw})F_{ST(ff),t-1} - F_{ST(vw),t-1}], \end{aligned} \quad (35)$$

$$\frac{1}{2N_{eV,t}} = \frac{q'_{mm}F_{ST(mm),t-1}}{4} + \frac{q'_{mf}F_{ST(mf),t-1}}{2} + \frac{q'_{ff}F_{ST(ff),t-1}}{4}. \quad (36)$$

Like the instantaneous effective sizes, the *F*-statistics change over the first few generations. However, in later generations, the various effective sizes attain the same asymptotic value while the *F*-statistics converge to their distinctive asymptotic values. The changes in *F*-statistics for the first 10 generations are shown in Figure 5 for a population with parameters $s = 20$, $N_m = 5$, $N_f = 50$, $d_m = 0.5$, $d_f = 0$, $\sigma^2_{mm} = 2$, $\sigma^2_{ff} = 1$, $\sigma^2_{mf} = 10$, $\sigma^2_{fm} = 0.1$, $\sigma_{m,mf} = 2$, and $\sigma_{f,mf} = 1$. The graphs are generated by numerical calculations by (20), (10), and (11). Because of the asymmetrical effects of the two sexes, the differences among the four fixation indexes, $F_{ST(mm),t}$, $F_{ST(mf),t}$, $F_{ST(ff),t}$, and $F_{ST,t}$, are evident in any generation. Similar to the monoecious case (WANG 1997), $F_{IS(vw)}$ reaches its equilibrium value in fewer generations than $F_{ST(vw)}$ and F_{IT}.

For incomplete subdivision, the instantaneous *F*-statistics will eventually attain their respective asymptotic values, $F_{IS(vw)}$, $F_{ST(vw)}$, and F_{IT}, while the instantaneous effective sizes $N_{eI,t}$, $N_{e\theta(vw),t}$, and $N_{eV,t}$ will reach the same asymptotic value, N_e. Thus in equilibrium, (34–36) reduce to

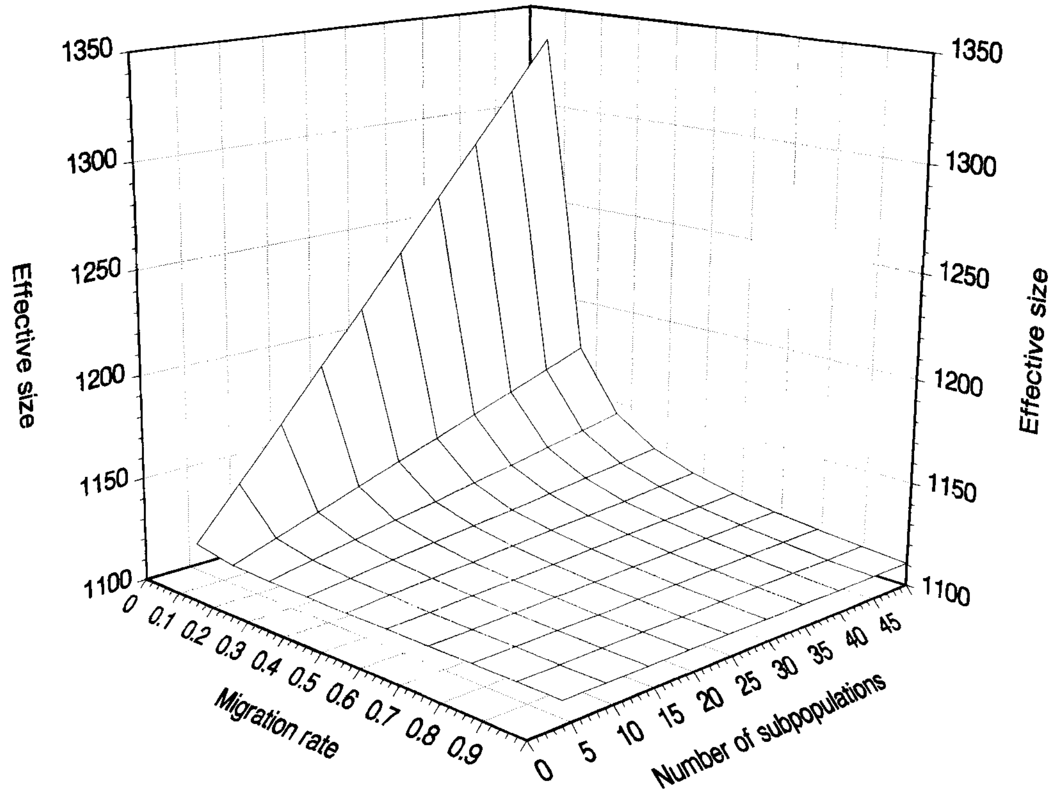

FIGURE 3.—Three dimensional diagrams showing the effect of the number of subpopulations (s) and migration rates ($d_m = d_f$) on the effective size (N_e). The graph plots equation 34 for a population with parameters $sN = 2000$, $N_f/N_m = 5$, $S^2_{mm} = S^2_{ff} = 1$, $S_{m,mf} = S_{f,mf} = 0$, and $S^2_{mf} = 1/S^2_{fm} = 5$.

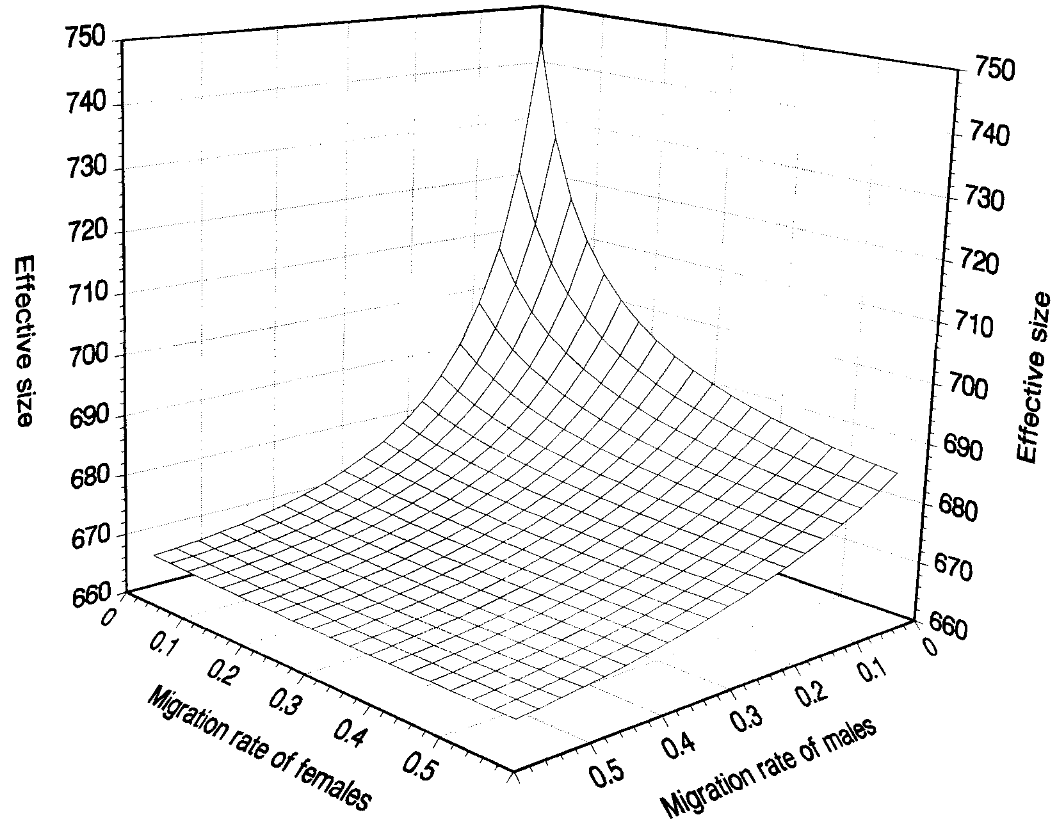

FIGURE 4.—Three dimensional diagrams depicting the influence of male and female migration rates (d_m, d_f) on the effective size (N_e). The graph is generated using 10 males and 50 females per subpopulation, 20 subpopulations and Poisson distribution of family size ($S^2_{mm} = S^2_{ff} = 1$, $S^2_{mf} = 1/S^2_{fm} = 5$, and $S_{m,mf} = S_{f,mf} = 0$).

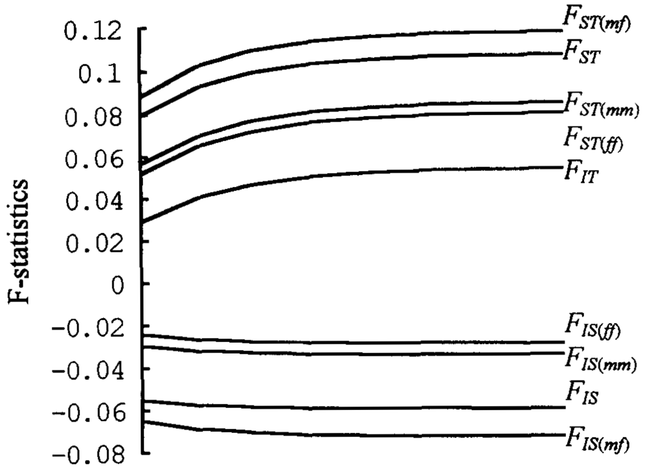

FIGURE 5.—Changes in the nine *F*-statistics parameters ($F_{IS(mm}$, $F_{IS(mf)}$, $F_{IS(ff)}$, F_{IS}, $F_{ST(mm)}$, $F_{ST(mf)}$, $F_{ST(ff)}$, F_{ST}, and F_{IT}) over successive generations for a population with parameters $s = 20$, $N_m = 5$, $N_f = 50$, $d_m = 0.5$, $d_f = 0$, $\sigma^2_{mm} = 2$, $\sigma^2_{ff} = 1$, $\sigma^2_{mf} = 10$, $\sigma^2_{fm} = 0.1$, $\sigma_{m,mf} = 2$, and $\sigma_{f,mf} = 1$.

$$\frac{1}{2N_e} = \frac{q_{mf}F_{ST(mf)} - F_{IT}}{1 - F_{IT}}, \tag{37}$$

$$\frac{1}{2N_e} = \frac{1}{8(1 - F_{ST(vw)})}[p_{vw}(1 + F_{IT}) + 2q_{mm}(1 - p_{m,vw})F_{ST(mm)} + 4q_{mf}F_{ST(mf)} + 2q_{ff}(1 - p_{f,vw})F_{ST(ff)} - F_{ST(vw)}], \tag{38}$$

$$\frac{1}{2N_e} = \frac{q'_{mm}F_{ST(mm)}}{4} + \frac{q'_{mf}F_{ST(mf)}}{2} + \frac{q'_{ff}F_{ST(ff)}}{4}. \tag{39}$$

A solution to these equations will give explicitly the expressions for F_{IT} and $F_{ST(vw)}$, and by (10) and the relation $(1 - F_{IT}) = (1 - F_{ST(vw)})(1 - F_{IS(vw)})$, we can obtain $F_{IS(vw)}$. The expressions for $F_{ST(vw)}$, however, are quite complicated and therefore not given here. For the special case that the numbers of male and female progeny per family are in independent Poisson distributions, we have $F_{ST(mm)} = F_{ST(mf)} = F_{ST(ff)} = F_{ST}$ because the coancestry between individuals within subpopulations is irrespective of the sexes of the individuals. Thus from (39) we get

$$F_{ST} = \frac{2}{(q'_{mm} + 2q'_{mf} + q'_{ff})N_e}. \tag{40}$$

Inserting (40) into (37) and (10) yields

$$F_{IT} = \frac{4q_{mf}/(q'_{mm} + 2q'_{mf} + q'_{ff}) - 1}{2N_e - 1} \tag{41}$$

and

$$F_{IS} = \frac{4q_{mf} - q'_{mm} - 2q'_{mf} - q'_{ff} - 4}{2(q'_{mm} + 2q'_{mf} + q'_{ff})N_e - 4} \tag{42}$$

approximately. For relatively large subpopulation size, we use (5) and (6) for simplicity and these equations reduce to

$$F_{ST} = \frac{2s}{(d_m + d_f)(4 - d_m - d_f)N_e}, \tag{43}$$

$$F_{IT} = \frac{4s(1 - d_m)(1 - d_f) + (d_m - d_f)^2}{(d_m + d_f)(4 - d_m - d_f)(2N_e - 1)}, \tag{44}$$

$$F_{IS} = \frac{4s(d_m + d_f - d_m d_f) - (d_m - d_f)^2}{4s - 2(d_m + d_f)(4 - d_m - d_f)N_e}. \tag{45}$$

For a single random mating population ($s = d_m = d_f = 1$), (45) simplifies to

$$F_{IS} = -\frac{1}{2N_e - 1} \tag{46}$$

which, for equal numbers of males and females ($N_m = N_f = N/2$) thus $N_e \approx N$, reduces to the classical result (KIMURA and CROW 1963; ROBERTSON 1965)

$$F_{IS} = -\frac{1}{2N - 1}. \tag{47}$$

For the special case that s is large, $N_m = N_f = N/2$ and both d_m and d_f are small enough so that second and higher order terms can be ignored, the effective size is

$$N_e = sN + \frac{s(1 - d_m - d_f)}{2(d_m + d_f)} \tag{48}$$

approximately from (32). Substituting (48) into (43) we obtain

$$F_{ST} = \frac{1}{2N(d_m + d_f) + 1}, \tag{49}$$

which was also derived by PROUT (1981). If we do not consider sex and denote migration rate as $m = \frac{1}{2}(d_m + d_f)$, (49) reduces to the classical result of WRIGHT (1969). Thus our expressions for *F*-statistics are general, incorporating a finite number of subpopulations, different numbers of males and females per subpopulation, sex-dependent migration rates, and an arbitrary distribution of family size.

For a subdivided population with equal numbers of males and females in each subpopulation and Poisson distribution of family size, CHESSER *et al.* (1993, Equation 51) obtained the estimate of F_{ST} as

$$F_{ST} = \frac{2s}{[3(d_m + d_f) - 2d_m d_f]N_e}. \tag{50}$$

The equation is clearly in variance with our expression 43, and for the infinite island model it also differs from the result by WRIGHT (1969) and PROUT (1981). Expression 50 gives correct predictions of F_{ST} only for complete migration ($d_m = d_f = 1$) or no migration ($d_m = d_f = 0$); otherwise it always overestimates F_{ST}, the largest bias being resulted with intermediate values of migration rates of male and female individuals. This is because (50) was derived from recurrence equations that used the incorrect probabilities (24) and (25).

DISCUSSION

The expressions for gene identity, effective size and *F*-statistics derived herein are applicable to subdivided dioecious populations exhibiting varying degrees of isolation and with different numbers of males and females in each subpopulation and an arbitrary distribution of family size. Traditional equations for a single unsubdivided population are shown to be special cases of our general expressions, and CHESSER and coworker's model is extended and also corrected in this study.

It is clear that population structure has an important effect on inbreeding and genetic drift in both short-term periods and in the long run. Migration among the subpopulations alleviates inbreeding for the first few generations, but it will result in a high final rate of inbreeding and thus larger inbreeding coefficients in later generations. These results can be used as a guide to determine proper values of rates of migration and subpopulation number and size in genetic conservation programs.

The effect of population structure on effective size is determined mainly by migration rates. The smaller the migration rate, the more generations are required for the population to attain its steady state in inbreeding and genetic drift (CHESSER *et al.* 1993), and the larger the asymptotic effective size. In instances of complete migration by one sex or both sexes, the population behaves as an unsubdivided random mating population and there is little difference between values of effective size predicted by traditional and the newly derived equations. When subpopulations become isolated the influence of breeding structure takes on greater importance in the determination of effective size (CHESSER *et al.* 1993). If there is no genetic exchange among subpopulations ($d_m = d_f = 0$), then inbreeding and variance effective sizes do not converge to the same value, and inbreeding coefficient increases and gene frequency changes with different asymptotic rates.

In our derivations, some simplifying assumptions are included, such as random mating within subpopulations, constant size and structure of the population, nonoverlapping generations and autosomal inheritance, and without-selection and mutation. HILL (1979) showed that, for a population of constant size and sex ratio and with a stable age structure distribution, the effective size is the same as that for a population with discrete generations having the same variance in lifetime progeny numbers and the same numbers of individuals entering the population in each generation. It is unlikely that substantial deviations from the numerical values will result if extensions of the expressions derived herein are made to include overlapping generations. Recently the effective size for a sex-linked locus has been considered by many authors for a single unsubdivided population (POLLAK 1990; CABALLERO 1994, 1995; NAGYLAKI 1995; WANG 1996). The methodologies provided here can also be extended to sex-linked loci. However, the situation becomes more complex because at least seven probabilities of identity by descent should be considered in the general model.

The author is grateful to two anonymous reviewers for their helpful comments on an earlier draft of this manuscript.

LITERATURE CITED

CABALLERO, A., 1994 Developments in the prediction of effective population size. Heredity **73:** 657–679.

CABALLERO, A., 1995 On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics **139:** 1007–1011.

CABALLERO, A., and W. G. HILL, 1992 Effective size of nonrandom mating populations. Genetics **130:** 909–916.

CHESSER, R. K., 1983 Genetic variability within and among populations of the black-tailed prairie dog. Evolution **37:** 320–331.

CHESSER, R. K., 1991a Gene diversity and female philopatry. Genetics **127:** 437–447.

CHESSER, R. K., 1991b Influence of gene flow and breeding tactics on gene diversity within populations. Genetics **129:** 573–583.

CHESSER, R. K., O. E. RHODES, JR., D. W. SUGG and A. F. SCHNABEL, 1993 Effective sizes for subdivided populations. Genetics **135:** 1221–1232.

FOLTZ, D. W., and J. L. HOOGLAND, 1983 Genetic evidence of outbreeding in the black-tailed prairie dog (*Cynomys ludovicianus*). Evolution **37:** 273–281.

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

KIMURA, M., and J. F. CROW, 1963 The measurement of effective population number. Evolution **17:** 279–288.

NAGYLAKI, T., 1995 The inbreeding effective population number in dioecious populations. Genetics **139:** 473–485.

NUNNEY, L., and D. R. ELAM, 1994 Estimating the effective size of conserved populations. Conserv. Biol. **8:** 175–184.

POLLAK, E., 1990 The effective population size of an age-structured population with a sex-linked locus. Math. Biosci. **101:** 121–130.

PROUT, T., 1981 A note on the island model with sex dependent migration. Theor. Appl. Genet. **59:** 327–332.

ROBERTSON, A., 1965 The interpretation of genotypic ratios in domestic animal populations. Anim. Prod. **7:** 319–324.

ROUX, C. Z., 1995 Countering inbreeding by migration. 2. Migration from related populations. South African J. Anim. Sci. **25:** 44–49.

SELANDER, R. K., 1970 Behavior and genetic variation in natural populations. Am. Zool. **10:** 53–66.

SUGG, D. W., and R. K. CHESSER, 1994 Effective population size with multiple paternity. Genetics **137:** 1147–1155.

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

WANG, J., 1997 Effective size and *F*-statistics of subdivided populations. I. Monoecious species with partial selfing. Genetics **146:** 000–000.

WAPLES, R. S., 1989 A generalized approach for estimating effective population size from changes in allele frequency. Genetics **121:** 379–391.

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

WRIGHT, S., 1969 *Evolution and the Genetics of Population*, Vol. 2. *The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

Communicating editor: B. S. WEIR