

GROUP INBREEDING AND COANCESTRY¹

C. CLARK COCKERHAM

*Department of Experimental Statistics, North Carolina State University,
Raleigh, North Carolina 27607*

Received November 30, 1966

IN his classical papers, "Systems of Mating" (1921), SEWALL WRIGHT developed the concept of the inbreeding coefficient in terms of the correlation of uniting gametes, and, among other things, related it to the decrease in heterozygosity and applied it to systems of mating relatives. The next year (1922) he produced procedures for calculating the inbreeding coefficient and a measure of the relationships among different individuals, which was now labeled coefficient of relationship, from pedigrees. In 1931 he showed the consequences of finite population size in terms of the inbreeding coefficient and the related variation and random drift of gene frequencies. In many other papers he and other workers have further refined the theory and extended the applications to more complex situations, but these three papers serve as the bases.

MALÉCOT (1948) provided an alternative view of the coefficient of inbreeding, making use of the probability of genes being "identical by descent," i.e., of being copies of the same ancestral gene. His concomitant measure of the relationships among individuals, called coancestry herein, is more simply related to the inbreeding coefficient than is the coefficient of relationship. While MALÉCOT's definitions and methods must lead to the same results as does WRIGHT's, they are generally easier to grasp and apply, requiring only simple probability arguments, for those not well versed in path coefficients.

The purpose of the present paper is to extend the definitions of the inbreeding and coancestry coefficients to include groups of individuals. With these extensions, and a few further definitions, one can work with pedigrees of groups of individuals or of subdivisions of a population with almost the same ease as one does with pedigrees of individuals. Here again, the classical results are obtained, but the ease of application of the methods allows one to consider situations which otherwise would appear to be formidable.

DEFINITIONS AND OPERATIONS

As will become clear, the definitions are for diploids for which segregation ratios are assumed to be normal and disturbances such as selection are assumed to be absent throughout.

For purposes of clarification, consider the pedigree of groups of individuals in

¹ Contribution from the Department of Experimental Statistics, North Carolina Agricultural Experiment Station, Raleigh. Published with the approval of the Director of Research as Paper No. 2304 of the Journal Series. This investigation was supported in part by Public Health Service Research Grant GM 11546.

Diagram 1. The upper case letters designate groups of individuals and the lower case letters sets of gametes.

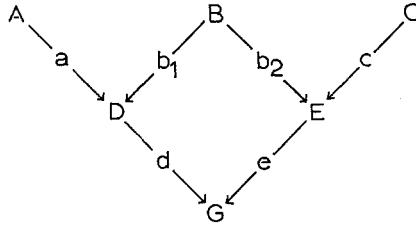


DIAGRAM 1

Corresponding to the inbreeding coefficient, F , of an individual let that, F_l , for a group of N individuals be the probability that a random pair of alleles are identical by descent. This amounts to an average of the $2N(2N-1)/2$ probabilities, one for each pairing of the $2N$ alleles. Of these pairs, N are within individuals and the remainder are among individuals. Averaging the probabilities of identity by descent, for the A group for example with N_A individuals, leads to

$$F_{lA} = \frac{\bar{F}_A}{2N_A - 1} + \frac{2(N_A - 1)}{2N_A - 1} \bar{\theta}_A = \frac{\bar{F}_A - \bar{\theta}_A}{2N_A - 1} + \bar{\theta}_A, \tag{1}$$

where \bar{F}_A is an average of the N_A inbreeding coefficients and $\bar{\theta}_A$ is an average of the $N_A(N_A - 1)/2$ coancestries of the individuals. Since \bar{F} is an expected value, applicable to a random member of the group, the bar will generally be dropped. The group or line inbreeding coefficient reduces to that of the individual when $N = 1$. For large lines, F_l is dominated by $\bar{\theta}$, the likeness among individuals or homogeneity of the group. The difference, $F - \bar{\theta}$, is indicative of the structuring of the group, i.e., a collection of individuals more or less related than they are inbred.

Corresponding to the coancestry of two individuals, θ , let that, θ_l , for two groups of individuals be the probability of a random allele from one group being identical by descent to a random allele from the other group. In the case of two distinct groups, A and B for example, it is just the average of the $N_A N_B$ coancestries of the A individuals with the B individuals,

$$\theta_{lAB} = \bar{\theta}_{AB}. \tag{2}$$

In the case of the coancestry of a group of individuals with itself, it is the average of the N^2 coancestries including individuals with themselves,

$$\theta_{lBB} = \frac{1 + F_B}{2N_B} + \frac{N_B - 1}{N_B} \bar{\theta}_B = \frac{1 + F_B - 2\bar{\theta}_B}{2N_B} + \bar{\theta}_B, \tag{3}$$

which reduces to that of an individual with itself when $N_B = 1$. Note also that θ_{lBB} is the probability of two random alleles with replacement being identical by descent. With probability $1/2N_B$ the two alleles will be copies of the same gene and identical by descent. With probability $(2N_B - 1)/2N_B$, they will be copies

of different genes but may still be identical by descent with probability F_{IB} . Putting these probabilities together,

$$\theta_{IBB} = \frac{1}{2N_B} + \frac{2N_B - 1}{2N_B} F_{IB}, \quad (4)$$

which is a different form of (3).

Independent sampling of gametic sets from the same group: For these definitions to be operationally useful, some rules concerning groups of individuals and gametic sets must be specified.

1. The probability that any individual of a parental group contributes a random gamete in an output set of gametes is the same for all members, $1/N$, and not dependent on gametes in other output sets from the same group.
2. Numbers of gametes in uniting input sets are equal, e.g., $N_a = N_{b_1} = N_D$.
3. Gametes from one input set unite at random with gametes from the other input set, i.e., mating is random.

With these rules we note that the probability of a random allele in one uniting set being identical by descent with a random allele in the other uniting set is the same as the inbreeding coefficient of the offspring group *and* the coancestry of the parental groups. For example,

$$P(d' \equiv e') = F_G = \theta_{IDE}, \quad (5)$$

where a prime is used to indicate a random gamete (allele) of a set. For the relationship between an incoming set of gametes and an output set of gametes for a group, we shall use

$$P(d' \epsilon a) = P(d' \epsilon b_1) = P(e' \epsilon b_2) = P(e' \epsilon c) = 1/2 \quad (6)$$

to denote the probability of a random gene in the output set having an ancestral gene in one of the input sets. Since each parental group contributes equally to the offspring group, this gamete to gamete transitional probability is $1/2$ as it is for individuals.

Expanding the argument of probability of identity by descent to more remote ancestral groups,

$$\begin{aligned} P(d' \equiv e') &= P(e' \epsilon b_2)P(d' \equiv b'_2) + P(e' \epsilon c)P(d' \equiv c') \\ &= 1/2 [P(d' \equiv b'_2) + P(d' \equiv c')] . \end{aligned} \quad (7)$$

Thus,

$$F_G = \theta_{IDE} = 1/2 [\theta_{IDB} + \theta_{IDC}] , \quad (8)$$

and the coancestry of one group with another is the average of the coancestries of one group with the parental groups of the other, just as it is for individuals. Further expanding,

$$P(d' \equiv b'_2) = 1/2 [P(a' \equiv b'_2) + P(b'_1 \equiv b'_2)] \quad (9)$$

$$P(d' \equiv c') = 1/2 [P(a' \equiv c') + P(b'_1 \equiv c')] ,$$

and

$$F_G = \theta_{IDE} = 1/4 [\theta_{IAB} + \theta_{IAC} + \theta_{IBB} + \theta_{IBC}] . \quad (10)$$

The coancestry of two groups is the average of the four coancestries between their two parental groups, which in this example includes the coancestry of one group, B, with itself. If groups A, B and C are unrelated, the inbreeding coefficient in (10) reduces to

$$F_G = \theta_{l_{BB}}/4. \quad (11)$$

Note that this is the result obtained by counting paths to common ancestral groups in the pedigree.

$$\begin{aligned} P(d' \equiv e') &= P(d' \epsilon b_1)P(b'_1 \equiv b'_2)P(e' \epsilon b_2) \\ &= P(b'_1 \equiv b'_2)/4 = \theta_{l_{BB}}/4 \end{aligned} \quad (12)$$

So far we have the same simplicity for groups as for individuals. However, with groups one must reckon also with $\bar{\theta}$ which is different from F for subdivided populations. The average coancestry among individuals within a group will involve genes from the two different uniting sets of gametes $1/2$ of the time and from each of the uniting sets $1/4$ of the time. Denoting the probability of two random alleles in a set of gametes being identical by descent as ψ ,

$$\bar{\theta}_G = \frac{\psi_d}{4} + \frac{\psi_e}{4} + \frac{F_G}{2} = \frac{\psi_d}{4} + \frac{\psi_e}{4} + \frac{\theta_{l_{DE}}}{2}. \quad (13)$$

We shall generally use a slightly different notation, ψ_{ij} (e.g., $\psi_{GD} = \psi_d$, $\psi_{GE} = \psi_e$), for the set of gametes received by group i from parental group j . The number of gametes in the set is N_i , the same as the number in the offspring group. Let the probability that two random gametes in a set come from the same parent be $1/N_{gj}$. Then they are from different parents with probability $(N_{gj}-1)/N_{gj}$. The probability of the alleles being identical by descent is $(1+F_j)/2$ in the first case and θ_j in the second case. Putting the probabilities together,

$$\psi_{ij} = \frac{1+F_j}{2N_{gj}} + \frac{N_{gj}-1}{N_{gj}} \bar{\theta}_j. \quad (14)$$

We shall relate N_{gj} , an effective number corresponding in idea to the effective population number (WRIGHT 1938), to the mean and variance of the number of gametes per parent. Let k_u be the number of gametes contributed by the u th parent. The total number of gametes is the same as the number of offspring,

$$\sum k_u = N_i, \quad (15)$$

with a mean of

$$\bar{k}_{ij} = \frac{\sum k_u}{N_j} = \frac{N_i}{N_j}, \quad (16)$$

which will be one if the parental and offspring groups are of the same size in contrast to 2 for the effective population number pertaining to a combined set of gametes. The total number of pairs of gametes is

$$N_i(N_i-1)/2 = N_j \bar{k}_{ij} (N_j \bar{k}_{ij} - 1)/2. \quad (17)$$

Of these,

$$\sum k_u(k_u-1)/2 = N_j [\sigma_{k_{ij}}^2 + \bar{k}_{ij}(\bar{k}_{ij}-1)]/2 \quad (18)$$

involve pairs from the same individual. Thus, the probability of two random alleles being from the same parent is

$$\frac{1}{N_{gj}} = \frac{\sigma_{k_{ij}}^2 + \bar{k}_{ij}(\bar{k}_{ij}-1)}{\bar{k}_{ij}(N_j\bar{k}_{ij}-1)}, \quad (19)$$

or the effective number is

$$N_{gj} = \frac{\bar{k}_{ij}(N_j\bar{k}_{ij}-1)}{\sigma_{k_{ij}}^2 + \bar{k}_{ij}(\bar{k}_{ij}-1)}, \quad (20)$$

which has the same expression as the effective population number for a monocious population (KIMURA and CROW 1963). When each parent has an equal chance, $1/N_j$, of contributing each gamete, the distribution of the number of gametes per parent is binomial,

$$f(k) = C_k^{N_i} \left(\frac{1}{N_j}\right)^k \left(\frac{N_j-1}{N_j}\right)^{N_i-k}, \quad (21)$$

with mean

$$\bar{k}_{ij} = \frac{N_i}{N_j}, \quad (22)$$

as before, and variance

$$\sigma_{k_{ij}}^2 = \bar{k}_{ij} \frac{N_j-1}{N_j}. \quad (23)$$

Substitution of (22) and (23) into (20) leads to $N_{gj} = N_j$, and the effective number is the census number of parents. In this case, with further substitution into (14) and comparing with (3), we have

$$\psi_{ij} = \frac{1+F_j}{2N_j} + \frac{N_j-1}{N_j} \bar{\theta}_j = \theta_{ijj}, \quad (24)$$

and the gametic probability is the same as the coancestry of the parental group with itself.

At the extreme of control of matings let each parent contribute the same number of gametes so that $\sigma_k^2 = 0$, which requires that \bar{k} take integral values. In this case,

$$N_{gj} = \frac{N_j\bar{k}_{ij}-1}{\bar{k}_{ij}-1}, \quad (25)$$

which has a minimum of N_j when \bar{k} is ∞ and a maximum of ∞ when $\bar{k} = 1$. Substitution of $N_{gj} = \infty$ into (14) leads to

$$\psi_{ij} = \bar{\theta}_j, \quad (26)$$

since no two gametes in a set come from the same parental individual. The same result (26) is obtained when $\bar{k} < 1$ and each contributing parent contributes only one gamete, because

$$\sigma_{k_{ij}}^2 = \left(N_i - \frac{N_i^2}{N_j}\right)/N_j = \bar{k}_{ij} - \bar{k}_{ij}^2, \quad (27)$$

and

$$N_{gj} = \infty, \psi_{ij} = \bar{\theta}_j. \quad (28)$$

Returning now to Diagram 1 and rewriting (13),

$$\bar{\theta}_G = 1/4(\psi_{GD} + \psi_{GE} + 2\theta_{IDE}), \quad (29)$$

we must specify the σ_k^2 's or corresponding N_g 's to further expand the relationships. In the case of equal chance of each parent in a group contributing each gamete in a set,

$$\psi_{GD} = \theta_{IDD}, \quad \psi_{GE} = \theta_{IEE}, \quad (30)$$

and

$$\bar{\theta}_G = 1/4(\theta_{IDD} + \theta_{IEE} + 2\theta_{IDE}). \quad (31)$$

On the other hand, if the groups are of equal size and $\sigma_k^2 = 0$,

$$\psi_{GD} = \bar{\theta}_D, \quad \psi_{GE} = \bar{\theta}_E, \quad (32)$$

and

$$\bar{\theta}_G = 1/4(\bar{\theta}_D + \bar{\theta}_E + 2\theta_{IDE}). \quad (33)$$

In each case the various measures can be expanded back to include ancestral groups.

Sampling of combined gametic sets from the same group: Certain situations require the treatment of combined gametic sets from one group or generation to another. To do so, consider the combined gametic output of a group; for example, the double set, $b_1 \cup b_2$, from B in Diagram 1. Reverting to our general notation, let i and i' denote the offspring groups in case they are different and j the parental group as before. The number of gametes in the combined set is $N_i + N_{i'} = N_{\bar{i}}$. Since dimensions were left general in developing the gametic probabilities (14) and effective numbers (20), we only introduce new notations, Ψ for ψ , N_e for N_g , $N_{\bar{i}}$ for N_i , $\bar{k}_{\bar{i}j}$ for \bar{k}_{ij} , and $\sigma_{\bar{k}_{\bar{i}j}}^2$ for $\sigma_{k_{ij}}^2$, to distinguish that we are treating the combined gametic output of group j . The mean number of gametes per parent is now

$$\bar{k}_{\bar{i}j} = \bar{k}_{ij} + \bar{k}_{i'j} = \frac{N_i + N_{i'}}{N_j}, \quad (34)$$

which, if i and i' are one and the same group, reduces to

$$\bar{k}_{\bar{i}j} = \frac{2N_i}{N_j}, \quad (35)$$

and in either case is 2 for a constant size of groups. The other measures for the combined set are

$$N_{e_j} = \frac{\bar{k}_{\bar{i}j}(N_j \bar{k}_{\bar{i}j} - 1)}{\sigma_{\bar{k}_{\bar{i}j}}^2 + \bar{k}_{\bar{i}j}(\bar{k}_{\bar{i}j} - 1)} \quad (36)$$

$$\Psi_{\bar{i}j} = \frac{1 + F_j}{2N_{e_j}} + \frac{N_{e_j} - 1}{N_{e_j}} \bar{\theta}_j. \quad (37)$$

Next consider the subdivision of the combined gametic set into the two sets of N_i and $N_{i'}$ gametes in a manner such that it is random with respect to the parents. We note now that the probability of alleles being identical by descent in two random gametes is the same for both gametic sets, $\Psi_{ij} = \Psi_{i'j} = \Psi_{\bar{i}j}$, and also is the same for a random gamete in one set and a random gamete from the other

set. For the gametic sets involving group B in Diagram 1, we have with this sampling plan,

$$[P(b'_1 \equiv b''_1) = \Psi_{DB}] = [P(b'_2 \equiv b''_2) = \Psi_{EB}] = [P(b'_1 \equiv b'_2) = \Psi_{BB}], \quad (38)$$

where primes indicate random genes (gametes).

A modification of rule one must be made for nonindependent sampling of gametic sets from the same group in order to operate with pedigrees of groups as we did previously. The transitional probabilities (6) are unaffected by this modification, and so is $\bar{\theta}$ (13),

$$\bar{\theta}_G = \frac{\Psi_d}{4} + \frac{\Psi_e}{4} + \frac{F_G}{2}, \quad (39)$$

except the gametic probabilities are for the appropriate sampling plan. The essential difference is that the coancestry of a population with itself, θ_{lBB} in (10, 11, 12), is replaced by Ψ_{BB} .

With equal chance of any parent contributing each gamete in a set, the effective number is unaffected by the number of gametes in the set, and

$$N_{ej} = N_{gj} = N_j. \quad (40)$$

Consequently,

$$\Psi_{ij} = \Psi_{i'j} = \Psi_{jj} = \theta_{ijj}, \quad (41)$$

and the results are the same as for independent sampling of gametic sets (24, 30). On the other hand, if all gametic variances are zero, which is compatible with $N_i = N_{i'} = N_j$ for example,

$$N_{ej} = 2N_j - 1, \quad (42)$$

and

$$\Psi_{ij} = \Psi_{i'j} = \Psi_{jj} = \frac{1 + F_j}{2(2N_j - 1)} + \frac{2(N_j - 1)}{2N_j - 1} \bar{\theta}_j \quad (43)$$

in comparison to

$$N_{gj} = \infty, \quad (44)$$

and

$$\psi_{ij} = \psi_{i'j} = \bar{\theta}_j \quad (45)$$

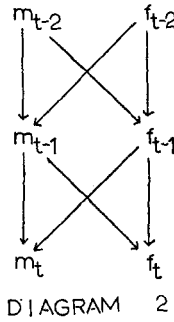
$$\theta_{ijj} = \frac{1 + F_j}{2N_j} + \frac{N_j - 1}{N_j} \bar{\theta}_j$$

for independently sampled sets of gametes. The two sampling plans are the same with equal chance of each parent contributing any gamete in a set and differences arise only when the gametic variance is different from that for equal chance.

With these modifications outlined for combined sampling of sets of gametes, pedigrees of groups can be treated in the same manner as for independent sampling of sets of gametes.

EXAMPLES

As an example, consider the following Diagram 2 for a dioecious population. The numbers in each sex will be assumed to be constant over time. Small m and f are used to designate male and female groups, respectively.



Using (10) and preceding formulae for independent sampling of sets of gametes,

$$F_t = F_{m_t} = F_{f_t} = \theta_{lmft-1} = \frac{1}{4}(\theta_{lmmt-2} + \theta_{lfft-2} + 2\theta_{lmft-2}) \tag{46}$$

$$= \frac{1}{4} \left[\left(\frac{1+F_{t-2}}{2N_m} + \frac{N_m-1}{N_m} \bar{\theta}_{mt-2} \right) + \left(\frac{1+F_{t-2}}{2N_f} + \frac{N_f-1}{N_f} \bar{\theta}_{ft-2} \right) + 2F_{t-1} \right]. \tag{47}$$

Turning now to the coancestry coefficients (13),

$$\begin{aligned} \bar{\theta}_{mt} &= \frac{1}{4}(\psi_{mmt-1} + \psi_{mft-1} + 2\theta_{lmft-1}) \\ \bar{\theta}_{ft} &= \frac{1}{4}(\psi_{ffft-1} + \psi_{fmt-1} + 2\theta_{lmft-1}), \end{aligned} \tag{48}$$

and we must specify the distribution of gametes per parent. With equal chance of each parent in a group contributing each gamete, we may use (24) and

$$\bar{\theta}_{mt} = \bar{\theta}_{ft} = \frac{1}{4}(\theta_{lmmt-1} + \theta_{lffft-1} + 2\theta_{lmft-1}), \tag{49}$$

which when compared to (46) is seen by inspection to be F_{t+1} . The substitution of F_{t-1} for $\bar{\theta}_{mt-2}$ and $\bar{\theta}_{ft-2}$ in (47) gives

$$F_t = F_{t-1} + \frac{1 + F_{t-2} - 2F_{t-1}}{2N_s} \tag{50}$$

$$\frac{1}{N_s} = \frac{1}{4N_m} + \frac{1}{4N_f},$$

the appropriate result obtained by WRIGHT (1931).

With a constant number for each sex over time $\bar{k}_{mm} = \bar{k}_{ff} = 1$, and one can control the gametes per parent so that $\sigma_{k_{mm}}^2 = \sigma_{k_{ff}}^2 = 0$, and the corresponding effective numbers are infinite (25). Further, for $N_m < N_f$, and each female contributing no more than one gamete to the male offspring group, the effective gametic number is also infinite (27, 28). Thus, from (26) and (28)

$$\psi_{mmt-1} = \bar{\theta}_{mt-1}, \psi_{mft-1} = \psi_{ffft-1} = \bar{\theta}_{ft-1}. \tag{51}$$

For integral values of $\bar{k}_{fm} > 1$, and with $\sigma_{k_{fm}}^2 = 0$, from (25) and (22) we find

$$N_{gm} = \frac{N_m(N_f-1)}{N_f-N_m}$$

$$\psi_{f_{mt-1}} = \frac{(N_f - N_m)}{N_m(N_f - 1)} \frac{(1 + F_{t-1})}{2} + \frac{N_f(N_m - 1)}{N_m(N_f - 1)} \bar{\theta}_{mt-1} \quad (52)$$

Substitution into (48) leads to different values for $\bar{\theta}_{mt}$ and $\bar{\theta}_{ft}$,

$$\begin{aligned} \bar{\theta}_{mt} &= 1/4 (\bar{\theta}_{mt-1} + \bar{\theta}_{ft-1} + 2F_t) \\ \bar{\theta}_{ft} &= 1/4 \left[\bar{\theta}_{ft-1} + \frac{(N_f - N_m)}{N_m(N_f - 1)} \frac{(1 + F_{t-1})}{2} + \frac{N_f(N_m - 1)}{N_m(N_f - 1)} \bar{\theta}_{mt-1} + 2F_t \right], \quad (53) \end{aligned}$$

unless $N_m = N_f$.

Consider first the situation in which $N_m = N_f = N/2$, and consequently,

$$\bar{\theta}_{mt} = \bar{\theta}_{ft} = \frac{F_t}{2} + \frac{\bar{\theta}_{mt-1}}{2} = \frac{F_t}{2} + \frac{F_{t-1}}{4} + \frac{F_{t-2}}{8} + \dots \quad (54)$$

Substitution of (54) into (47) gives for the recurrence relation of F ,

$$F_t = \frac{1 + F_{t-2}}{2N} + \frac{F_{t-1}}{2} + \frac{N-2}{2N} \left(\frac{F_{t-2}}{2} + \frac{F_{t-3}}{4} + \frac{F_{t-4}}{8} + \dots \right) \quad (55)$$

Adding $F_{t-1}/2$ to the right hand side of (55) and subtracting it out in the corresponding expression for F_t leads to

$$F_t = F_{t-1} + \frac{1 - F_{t-3}}{4N}, \quad (56)$$

which is the result given by ROBINSON and BRAY (1965) for a slightly different but analogous situation. If the initial (zeroth generation) members are noninbred, inbreeding is avoided in the first generation and is $1/2N$ in the second generation, after which the inbreeding coefficient may be found by repeated applications of (56). The average coancestry within a group may be found by either from of (54) and begins with $1/4N$ in the second generation. The average coancestry between males and females is $\theta_{lmft} = F_{t+1}$ for all variations in numbers of each sex and of ψ 's.

For unequal numbers in each sex, the statistics for the first three generations are

Generation	F	$\bar{\theta}_m$	$\bar{\theta}_f$
0	0	0	0
1	0	0	$\frac{N_f - N_m}{8N_m(N_f - 1)}$ (57)
2	$\frac{1}{2N_s}$	$\frac{1}{4N_s} + \frac{N_f - N_m}{32N_m(N_f - 1)}$	$\frac{1}{4N_s} + \frac{5(N_f - N_m)}{32N_m(N_f - 1)}$

where N_s was defined in (50). These statistics may be carried forward by repeated applications of (53) and (47). An alternative recursion form, similar in some respects to (56), still requires an accounting of the $\bar{\theta}$'s.

$$F_t = F_{t-1} + \left(\frac{3}{32N_m} + \frac{1}{32N_f} \right) (1 - F_{t-3}) \quad (58)$$

$$+ \frac{N_f - N_m}{16N_m N_f} (F_{t-3} - \bar{\theta}_{ft-3}).$$

The last term vanishes with $N_f = N_m$ and the result is the same as given in (56). The last term is small with large N_m and N_f because of the coefficient and because of the concomitant effect on the difference $F_{t-3} - \bar{\theta}_{ft-3}$. The difference also decreases with time since F and $\bar{\theta}$ converge, so that with good approximation in many cases the last term can be dropped. The term $\left(\frac{3}{32N_m} + \frac{1}{32N_f}\right)$ was given by GOWE, ROBERTSON and LATTER (1959) as the reciprocal of twice the effective population number for the drift variance.

If we consider now that sex of the offspring is determined at random for any mating, subject only to the restriction that N_m male and N_f female offspring are produced, the gametic sampling plan is that outlined for combined gametic sets from a parental group. A male (female) parent may have only male (female) offspring, the outcome being determined randomly.

With the modifications (34 through 39) for nonindependent sampling of gametic sets,

$$F_t = F_{mt} = F_{ft} = \theta_{lmft-1} \\ = 1/4 (\Psi_{mmt-2} + \Psi_{ffft-2} + 2\theta_{lmft-2}) \quad (59)$$

$$= 1/4 \left[\left(\frac{1+F_{t-2}}{2N_{em}} + \frac{N_{em}-1}{N_{em}} \bar{\theta}_{mt-2} \right) + \left(\frac{1+F_{t-2}}{2N_{ef}} + \frac{N_{ef}-1}{N_{ef}} \bar{\theta}_{ft-2} \right) + 2F_{t-1} \right]. \quad (60)$$

The average coancestries of males and of females are always the same,

$$\bar{\theta}_{mt} = \bar{\theta}_{ft} = 1/4 (\Psi_{mmt-1} + \Psi_{ffft-1} + 2\theta_{lmft-1}), \quad (61)$$

and when compared to (59) is seen to be

$$\bar{\theta}_{mt} = \bar{\theta}_{ft} = \theta_{lmft} = F_{t+1}. \quad (62)$$

Substitution of F_{t-1} for $\bar{\theta}_{mt-2}$ and $\bar{\theta}_{ft-2}$ into (60) leads to a general recurrence form for F ,

$$F_t = F_{t-1} + \frac{1 + F_{t-2} - 2F_{t-1}}{2} \left(\frac{1}{4N_{em}} + \frac{1}{4N_{ef}} \right) \\ = F_{t-1} + \frac{1 + F_{t-2} - 2F_{t-1}}{2N_e}. \quad (63)$$

Writing the means and effective numbers (34, 36) for males in notations appropriate for this example, we have

$$\bar{k}_{mm} = \bar{k}_{mm} + \bar{k}_{fm} = \frac{N_m + N_f}{N_m} = \frac{N}{N_m} \quad (64)$$

$$N_{em} = \frac{\bar{k}_{mm} (N_m \bar{k}_{mm} - 1)}{\sigma_{k_{mm}}^2 + \bar{k}_{mm} (\bar{k}_{mm} - 1)} = \frac{N_m N (N - 1)}{N_m^2 \sigma_{k_{mm}}^2 + N N_f}, \quad (65)$$

and the corresponding ones for females are obtained by interchanging m and f .

The overall effective number in (63),

$$N_e = \frac{4N_{em}N_{ef}}{N_{em} + N_{ef}}, \quad (66)$$

although not readily apparent, is the same as the one given by KIMURA and CROW (1963) for separate sexes, and it was this sampling plan that they treated. For equal chance of parents, $N_{em} = N_m$, $N_{ef} = N_f$, $N_e = N_s$, and the result (63) is the same as in (50), which we already knew from (40, 41). For equal numbers of each sex and equal gametic variances, $N_e = 2N_{em}$ and has a maximum value of $N_e = 2(N-1)$ with zero gametic variance. For unequal numbers in each sex, $N_m < N_f$, and for integral values of \bar{k}_{fm} , the gametic variance for males can be made zero, and from (65)

$$N_{em} = \bar{k}_{mf} (N-1) = \frac{N_m(N-1)}{N_f} . \tag{67}$$

The minimum gametic variance for females, however, is

$$\sigma_{\bar{k}_{ff}}^2 = \bar{k}_{mf}(1-\bar{k}_{mf}) = \frac{N_m(N_f-N_m)}{N_f^2} . \tag{68}$$

Substitution of (68) into the counterpart of (65) for females gives

$$N_{ef} = \left(\frac{\bar{k}_{mm} + \bar{k}_{fm}}{2} \right) (N-1) = \frac{N(N-1)}{2N_m} . \tag{69}$$

Putting (67) and (69) together for an overall effective number, we have

$$N_e = \frac{4N_m N(N-1)}{N^2 + N_m(N_m - N_f)} . \tag{70}$$

Inbreeding starts in the second generation with a value of $1/2N_e$ and accrues according to (63). While the initial inbreeding is less, the rate of inbreeding is slightly greater than for independent sampling of gametic sets, except with equal chance of the parents when the two sampling methods are the same. Another difference, except for equal chance, is that always $\bar{\theta}_{mt} = \bar{\theta}_{ft} = \theta_{imft}$ for the combined gametic sampling, whereas all three may be different with independent sampling of gametic sets.

Consider as a final example a monoecious population, Diagram 3. With sexuality of gametes, gametic sets are involved, and the numbers of each, pollen and egg cells, must be the same, just as they are for separate sexes.



DIAGRAM 3.

For independent sampling of the two sets of gametes, and dropping the lettered subscripts,

$$F_t = \theta_{tt-1} = \frac{1+F_{t-1}}{2N} + \frac{N-1}{N} \bar{\theta}_{t-1} , \tag{71}$$

and from (29),

$$\bar{\theta}_{t-1} = \frac{\psi_{t-2}}{2} + \frac{\theta_{t-2}}{2}, \quad (72)$$

since the two gametic probabilities are the same. With equal chance of each parent contributing any gamete in the set,

$$\psi_{t-1} = \theta_{t-1} = F_t, \quad (73)$$

and

$$\bar{\theta}_t = F_t \text{ or } \bar{\theta}_{t-1} = F_{t-1}, \quad (74)$$

so that

$$F_t = F_{t-1} + \frac{1-F_{t-1}}{2N} \quad (75)$$

given by WRIGHT (1931).

If each parent furnishes one male and one female gamete, then $\sigma_k^2 = 0$ for both gametic sets. In this case,

$$\psi_t = \bar{\theta}_t, \quad (76)$$

and

$$\bar{\theta}_t = \frac{F_t}{2} + \frac{\bar{\theta}_{t-1}}{2} = \frac{F_t}{2} + \frac{F_{t-1}}{4} + \frac{F_{t-2}}{8} + \dots, \quad (77)$$

which leads to

$$F_t = \frac{1+F_{t-1}}{2N} + \frac{N-1}{N} \left(\frac{F_{t-1}}{2} + \frac{F_{t-2}}{4} + \dots \right). \quad (78)$$

Proceeding in the same manner as from (55) to (56),

$$F_t = F_{t-1} + \frac{1-F_{t-2}}{4N}. \quad (79)$$

The following general formula encompasses both (75) and (79) as special cases,

$$F_t = F_{t-1} + \frac{1-F_{t-2}}{4N} + \frac{1+F_{t-2}-2F_{t-1}}{4N_g}. \quad (80)$$

The inbreeding coefficient is $1/2N$ in the first generation and proceeds with time according to (80).

If male and female gametes are random with respect to the parents, we turn to combined gametic sampling, and find

$$F_t = \Psi_{t-1} = \frac{1+F_{t-1}}{2N_e} + \frac{N_e-1}{N_e} \bar{\theta}_{t-1}, \quad (81)$$

and

$$\bar{\theta}_{t-1} = \frac{F_{t-1}}{2} + \frac{\Psi_{t-2}}{2} = F_{t-1}. \quad (82)$$

The recurrence formula for F , and effective number, N_e ,

$$F_t = F_{t-1} + \frac{1-F_{t-1}}{2N_e}, \quad (83)$$

are those given by KIMURA and CROW (1963). The effective number, N_e , is $2(2N-1)/(\sigma_k^2+2)$ for a constant size of population (WRIGHT 1938), with a maxi-

mum of $2N-1$ when $\sigma_k^2 = 0$. Inbreeding is $1/2N_e$ in the first generation and increases with time according to (83). Except for equal chance when the two gametic sampling methods are the same, inbreeding is less initially but increases at a faster rate than for independent sampling of the two gametic sets. Also the probabilities of genes being identical by descent are always the same within and among individuals, i.e., $F_t = \bar{\theta}_t$, whereas they are not the same for independent sampling of gametic sets and reduced gametic variance.

When self-fertilization is avoided in a monoecious population, the differences that arise between the two methods of specifying gametic samples disappear. With independent sampling of the two sets of gametes, the average coancestry is

$$\bar{\theta}_t = \frac{\psi_{t-1}}{2} + \frac{1}{2} \left[\frac{1+F_{t-1}}{2(N-1)} + \frac{2N-3}{2(N-1)} \bar{\theta}_{t-1} \right] \quad (84)$$

$$\begin{aligned} &= \bar{\theta}_{t-1} + \frac{1+F_{t-1}-2\bar{\theta}_{t-1}}{2} \left(\frac{1}{2N_g} + \frac{1}{2(N-1)} \right) \quad (85) \\ &= F_{t+1} \end{aligned}$$

for the avoidance of self-fertilization. The term in the brackets (84) replaces θ_{t-1} in (72) because gametes are not united at random. Of the $N(N-1)$ pairs of genes among individuals which also involve genes between the two gametic sets, all of those pairs from the same parent must be among individuals when self-fertilization is not allowed. The expected number of pairs from the same parent is N regardless of N_g , and genes in the remaining pairs, $N(N-2)$, are related as among random parents. The effective number in (85),

$$N_e = \frac{2N_g(N-1)}{N_g+N-1} = \frac{4(N-1)}{2+2\sigma_k^2}, \quad (86)$$

is found to be equivalent to the one for the combined gametic set for which the gametic variance is twice that, σ_k^2 , of a half set. Since $F_t = \bar{\theta}_{t-1}$ in both situations, the recursion formulas for F and $\bar{\theta}$ are the same for both types of gametic sampling, and equivalent to those for separate sexes, equal numbers in each sex and sex of offspring determined randomly.

DISCUSSION

Many of the applications and simplifications which stem from the definitions and concepts developed herein for groups or subdivisions of populations remain to be elaborated. Some require further exploration. In particular, generalizations concerning avoidance and rate of inbreeding and the variance of gene frequencies due to random drift will be published subsequently.

The line inbreeding and coancestry coefficients, and their component parts, provide different measures of probabilities of identity by descent of genes among and within subdivisions and within individuals. While primary emphasis has been placed on the manipulations involving subdivisions of a population, the same concepts may be applied to the total finite population. It turns out in this connection that the coancestry of the population with itself, θ_{I^*} , bears a simple rela-

tionship to F in certain situations whereas the average coancestry, $\bar{\theta}_p$, does in other situations.

With separate sexes, and sex of offspring random with respect to the parents, we have from (62),

$$\bar{\theta}_{mt} = \bar{\theta}_{ft} = \theta_{lmft} = F_{t+1}. \quad (87)$$

Thus,

$$\bar{\theta}_{pt} = F_{t+1} \quad (88)$$

for all variations in gametic variances, and $\bar{\theta}_p$ proceeds as F just one generation in advance.

Giving equal weight to each sex since their gametic contributions are the same, the population coancestry coefficient is

$$\theta_{lpt} = \frac{1}{4}[\theta_{lmmt} + \theta_{lfft} + 2\theta_{lmft}]. \quad (89)$$

When gametic sets are independently sampled, from (46), the population coancestry coefficient,

$$\theta_{lpt} = F_{t+2},$$

proceeds the same as F two generations in advance. With equal chance, both relationships hold,

$$\theta_{lpt} = \bar{\theta}_{pt+1} = F_{t+2}. \quad (90)$$

The same kind of situation is found for monoecious populations. If type of gamete is random with respect to the parent, from (82),

$$\bar{\theta}_{pt} = \bar{\theta}_t = F_t, \quad (91)$$

and genes are related the same within and among individuals. On the other hand, the population coancestry coefficient is related to F simply for independently sampled sets of male and female gametes. From (71),

$$\theta_{lpt} = \theta_{lt} = F_{t+1}. \quad (92)$$

Again, with equal chance, both relationships hold,

$$\theta_{lpt} = \bar{\theta}_{t+1} = F_{t+1}. \quad (93)$$

The line or group coancestry coefficient plays an analogous role in the average relationship of relatives stemming from parental lines as the coancestry coefficient does for relatives from individual parents. This should be obvious from the definitions but may need emphasizing in terms of the covariances of relatives. Referring to Diagram 1 as a pedigree of individuals, D and E are half sibs. If A , B and C are unrelated,

$$\theta_{DE} = \frac{\theta_{BB}}{4} = \frac{1+F_B}{8}. \quad (94)$$

As a pedigree of groups, D and E are half-sib groups. Relationships of individuals in D with those in E can range from true half sibs to lesser related relatives. Again, assuming parental groups A , B and C to be unrelated, the average coancestry of relatives between the half sib groups is expected to be

$$\theta_{lDE} = \frac{\theta_{lBB}}{4} = \frac{1+F_B}{8N_B} + \frac{N_B-1}{4N_B} \bar{\theta}_B, \quad (95)$$

which reduces to the correct result for an individual parent, $N_B = 1$.

Under certain assumptions, COCKERHAM (1963), when individual mates are unrelated, the covariances of relatives may be expressed in the following form

$$C = \alpha\sigma_A^2 + \delta\sigma_D^2 + \alpha^2\sigma_{AA}^2 + \alpha\delta\sigma_{AD}^2 + \dots, \quad (96)$$

where the variances are the additive, dominance, and so on components indicated by their subscripts. The coefficients, α and δ , depend upon the relationship of the relatives such that $\alpha = 2\theta$, or is twice the coancestry of the two relatives, and δ is the probability that both alleles of one relative are identical by descent to both alleles of the other relative. For certain relatives which are related only through their paternal (m) genes and/or through their maternal (f) genes, MALÉCOT (1948) showed that

$$\alpha = \frac{\phi_m + \phi_f}{2}, \quad \delta = \phi_m\phi_f, \quad (97)$$

where $\phi_m(\phi_f)$ is the probability of the paternal (maternal) genes being identical by descent. For half and full sib relatives,

$$\phi_m = \theta_{mm} = \frac{1+F_m}{2}, \quad \phi_f = \theta_{ff} = \frac{1+F_f}{2}, \quad (98)$$

where $\theta_{mm}(\theta_{ff})$ is the coancestry of the paternal (maternal) parent with itself.

It should be clear by now that if parental lines replace individual parents, one need only replace θ with θ_l . For paternal half sib groups from paternal lines,

$$\phi_m = \theta_{lmm}, \quad \phi_f = 0, \quad \alpha = \frac{\theta_{lmm}}{2}, \quad \delta = 0. \quad (99)$$

In this case, θ_{lmm} is the coancestry of a paternal line with itself corresponding to θ_{mm} in (98). Similarly, for maternal half sib groups,

$$\phi_m = 0, \quad \phi_f = \theta_{lff}, \quad \alpha = \frac{\theta_{lff}}{2}, \quad \delta = 0, \quad (100)$$

and for full sib groups,

$$\phi_m = \theta_{lmm}, \quad \phi_f = \theta_{lff}, \quad \alpha = \frac{\theta_{lmm} + \theta_{lff}}{2}, \quad \delta = \theta_{lmm}\theta_{lff}. \quad (101)$$

When the parental lines are homozygous, $F = \bar{\theta} = 1$, and the result is the same as for a homozygous individual which it should be.

Mating designs of lines may be accomplished in the same manner as they are for individuals (СОКЕРНАМ 1963). In some species certain designs such as the factorial and diallel can be accommodated only by using lines. In cases where lines have been used in the estimation of genetic variances, they have been assumed to be homozygous. Now, one can make valid interpretations of the design components of variance in terms of covariances of relatives and genetic variances for partially inbred lines, providing that the line coancestries are known sufficiently well. The less homozygous and related are the members of the parental lines, the smaller are the coefficients, α and δ , however, which reduces the precision of any estimates of genetic variances.

SUMMARY

Definitions analogous to the inbreeding coefficient and the coancestry coefficient of individuals, in terms of the probability of genes being "identical by descent," were developed for groups of individuals. Additional probability meas-

ures of genes being identical by descent in gametic sets were defined and related to the other probability measures. For two methods of sampling, independent and combined, of sets of gametes from the same group, but always with random union (mating) of gametes between uniting sets, procedures were developed for operating with pedigrees of groups of individuals, or subdivisions of populations, in a manner analogous, although with some more complications, to that of pedigrees of individuals.—The various probability measures taken together provide a quantitative accounting of the structuring of a population. For any system of mating continued over time, recursion formulae for each measure can be established. Examples are given for dioecious and monoecious populations with variations not considered previously.—It is also pointed out how the coancestries of parental lines with themselves play a role corresponding to coancestries of parental individuals in the estimation of genetic variances from a mating design utilizing lines instead of individuals.

LITERATURE CITED

- COCKERHAM, C. C., 1963 Estimation of genetic variances. *Statistical Genetics and Plant Breeding*. Natl. Acad. Sci.-Natl. Res. Council Publ. **982**: 53–94.
- GOWE, R. S., A. ROBERTSON, and B. D. H. LATTER, 1959 Environment and poultry breeding problems. 5. The design of poultry control strains. *Poultry Science* **38**: 462–471.
- KIMURA, M., and J. F. CROW, 1963 The measurement of effective population number. *Evolution* **17**: 279–288.
- MALÉCOT, G., 1948 *Les Mathématiques de l'Hérédité*. Masson, Paris.
- ROBINSON, P., and D. F. BRAY, 1965 Expected effects on the inbreeding coefficient and rate of gene loss of four methods of reproducing finite diploid populations. *Biometrics* **21**: 447–458.
- WRIGHT, S., 1921 Systems of mating. *Genetics* **6**: 111–178. — 1922 Coefficients of inbreeding and relationship. *Am. Naturalist* **56**: 330–338. — 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159. — 1938 Size of population and breeding structure in relation to evolution. *Science* **87**: 430–431.