

THE INBREEDING EFFECTIVE POPULATION NUMBER
AND THE EXPECTED HOMOZYGOSITY
FOR AN X-LINKED LOCUS†

THOMAS NAGYLAKI

*Department of Biophysics and Theoretical Biology, The University of Chicago,
920 East 58th Street, Chicago, Illinois 60637*

Manuscript received November 2, 1980

ABSTRACT

Assuming random mating and discrete nonoverlapping generations, the inbreeding effective population number, $N_e^{(i)}$, is calculated for an X-linked locus. For large populations, the result agrees with the variance effective population number. As an application, the maintenance of genetic variability by the joint action of mutation and random drift is investigated. It is shown that, if every allele mutates at rate u to new types, then the probabilities of identity in state (and hence the expected homozygosity of females) converge to the approximate value $(1 + 4N_e^{(i)}u)^{-1}$ at the approximate asymptotic rate $\exp\{-[2u + (2N_e^{(i)})^{-1}]t\}$.

THE simplest classical model of inbreeding in a finite, panmictic population (WRIGHT 1931; MALÉCOT 1946, 1948; KIMURA 1963) posits that the organism is monoecious and that gametes are sampled with replacement. The introduction of an inbreeding effective population number allows the removal of these restrictive assumptions.

The generalization to separate sexes is fairly straightforward: consult WRIGHT (1931), MALÉCOT (1946, 1948) and KIMURA (1963) for the analysis of the autosomal case and WRIGHT (1933), MALÉCOT (1951) and KIMURA (1963) for the treatment of the X-linked one.

Sampling with replacement corresponds to the production of an extremely large gametic pool, from which a much smaller number of successful gametes survive to form the next generation. This process clearly implies that the number of successful gametes contributed by a given individual to the next generation has a binomial distribution. However, even in species (such as many plants, insects, and fish) that produce gametes greatly in excess of the number of surviving individuals, sundry random factors will usually raise or lower the contribution of some individuals to the large gametic pool, thereby increasing the variance of the number of successful gametes contributed by a particular individual. Indeed, the gametic variance observed in natural and laboratory populations, including man, exceeds the binomial value (CROW and MORTON 1955). The inbreeding effective population number with an arbitrary distribution of the

† Supported by National Science Foundation Grant No. DEB77-21494.

number of progeny has been studied extensively for an autosomal locus: refer to WRIGHT (1938, 1939), CROW (1954), KIMURA and CROW (1963), CROW and KIMURA (1970, pp. 345-352), CHIA and POLLAK (1974) and POLLAK (1977).

Since *X*-linked recessives are expressed in males, *X*-linked loci are of particular interest and usefulness in human, mouse and *Drosophila* genetics. Furthermore, arrhenotoky occurs in the insect orders Hymenoptera, Coleoptera, Thysanoptera and Homoptera, the arachnid order Acarina, and the rotifer order Monogonanta (HARTL 1971); in such animals, all loci are effectively *X*-linked.

In this note, we shall derive the inbreeding effective population number for an *X*-linked locus with an arbitrary progeny-size distribution. We shall apply the analysis to the calculation of the equilibrium values and the ultimate rate of convergence of the probabilities of identity (and hence the expected homozygosity of females) in the infinite alleles model (MALÉCOT 1946, 1948; WRIGHT 1949; KIMURA and CROW 1964) for a sex-linked locus.

THE INBREEDING EFFECTIVE POPULATION NUMBER

We assume that N_1 males and N_2 females mate at random; generations are discrete and nonoverlapping. The number of alleles at the *X*-linked locus under consideration is arbitrary; selection and mutation are absent. Let P_1 (P_2) denote the constant probability that two distinct, homologous, randomly chosen paternal (maternal) genes in generation t ($= 0, 1, 2, \dots$) come from the same individual in generation $t-1$. Designate by f_t the probability that a female chosen at random in generation t is homozygous; g_t signifies the probability that two homologous genes, one chosen at random from a male and one from a female, are the same allele. Let k_t represent the probability that two homologous genes, chosen at random from distinct males, are the same allele; l_t is the corresponding probability for two females. At least in principle, these probabilities of identity in state are measurable in natural populations.

The recursion relations read

$$f' = g, \quad (1a)$$

$$g' = \frac{1}{2}g + \frac{1}{2}[P_2 \frac{1}{2}(1+f) + (1-P_2)l], \quad (1b)$$

$$k' = P_2 \frac{1}{2}(1+f) + (1-P_2)l, \quad (1c)$$

$$l' = \frac{1}{4}[P_1 + (1-P_1)k] + \frac{1}{4}[P_2 \frac{1}{2}(1+f) + (1-P_2)l] + \frac{1}{2}g, \quad (1d)$$

where the prime indicates the next generation. Equation (1a) follows directly from the definitions of f and g . To obtain (1b), note that a gene in a male is necessarily derived from a female, while one in a female is equally likely to have come from a male or a female. In the first case, the probability of identity in state in the previous generation is just g ; in the second, it is $\frac{1}{2}(1+f)$ if the two maternal genes come from the same female (with probability P_2) and l if they

are derived from different females (with probability $1-P_2$). Similar reasoning yields (1c) and (1d).

Define the probabilities of nonidentity:

$$h = 1 - f, \quad x = 1 - g, \quad y = 1 - k, \quad z = 1 - l. \tag{2}$$

The probability that a randomly chosen female is heterozygous and the expected heterozygosity of females are both equal to h . From (1) and (2), we deduce the matrix difference equation $\mathbf{V}' = B\mathbf{V}$, where

$$\mathbf{V} = \begin{pmatrix} h \\ x \\ y \\ z \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{4}P_2 & \frac{1}{2} & 0 & \frac{1}{2}(1-P_2) \\ \frac{1}{2}P_2 & 0 & 0 & 1-P_2 \\ \frac{1}{8}P_2 & \frac{1}{2} & \frac{1}{4}(1-P_1) & \frac{1}{4}(1-P_2) \end{pmatrix}.$$

The eigenvalues of B are zero and the three roots of

$$8\lambda^3 - 2(2+a)\lambda^2 - (2-a+2b)\lambda + b = 0, \tag{3a}$$

where

$$a = 1 - P_2, \quad b = (1 - P_1)(1 - P_2). \tag{3b}$$

In a population of N monoecious individuals that reproduce by sampling with replacement, the probability of heterozygosity decays at the rate $[1 - (2N)^{-1}]^t$ (WRIGHT 1931; MALÉCOT 1946, 1948; KIMURA 1963). Since, in situations of evolutionary interest, usually many individuals reproduce (WRIGHT 1978, Ch. 2), we suppose that $P_1, P_2 \ll 1$. Then we may define the inbreeding effective population number, $N_e^{(i)}$, by writing the dominant root of (3) as

$$\lambda_0 = 1 - (2N_e^{(i)})^{-1} + O(P_1^2 + P_2^2) \tag{4}$$

as $P_1 \rightarrow 0$ and $P_2 \rightarrow 0$. Substituting (4) into (3) and expanding, we find

$$N_e^{(i)} = \frac{9}{2(P_1 + 2P_2)}. \tag{5}$$

It remains to calculate P_1 and P_2 in terms of the means and variances of the gametic distributions. Denote by $K_{s,j}$ the (random) number of successful gametes contributed by the j th ($j = 1, 2, \dots, N_s$) individual with sex s ($= 1, 2$), and introduce

$$\mu_s = E(K_{s,j}), \quad \sigma_s^2 = \text{Var}(K_{s,j}). \tag{6}$$

Taking the expectation of

$$\sum_{j=1}^{N_s} K_{s,j} = N,$$

in which $N = N_1 + N_2$ is the total population number, yields

$$\mu_s = N/N_s . \tag{7}$$

Given the gametic distribution $\mathbf{K}_s = \mathbf{k}_s$, two distinct, homologous gametes from a parent of sex s may be chosen in

$$\sum_{j=1}^{N_s} \binom{k_{s,j}}{2}$$

ways. Dividing this by the total number of choices of two gametes from individuals of the same sex, $\binom{N}{2}$, gives the conditional probability that two gametes are derived from the same parent of sex s . Therefore,

$$\begin{aligned} P_s &= \sum_{\mathbf{k}_s} P(\mathbf{K}_s = \mathbf{k}_s) \frac{\sum_{j=1}^{N_s} \binom{k_{s,j}}{2}}{\binom{N}{2}} \\ &= [N(N-1)]^{-1} \sum_{\mathbf{k}_s} P(\mathbf{K}_s = \mathbf{k}_s) \sum_{j=1}^{N_s} k_{s,j}(k_{s,j} - 1) \\ &= [N(N-1)]^{-1} E \left[\sum_{j=1}^{N_s} K_{s,j}(K_{s,j} - 1) \right] \\ &= \frac{N_s(\sigma_s^2 + \mu_s^2 - \mu_s)}{N(N-1)} . \end{aligned} \tag{8}$$

Inserting (8) into (5), we obtain

$$N_e^{(i)} = \frac{9N(N-1)}{2[N_1(\sigma_1^2 + \mu_1^2 - \mu_1) + 2N_2(\sigma_2^2 + \mu_2^2 - \mu_2)]} , \tag{9}$$

with $\mu_s = N/N_s$.

For $N \gg 1$, (9) agrees with the asymptotic expression for the variance effective population number derived by MORAN and WATTERSON (1959) for two alleles and ETHIER and NAGYLAKEI (1980) for many.

If $K_{s,j}$ is binomially distributed with index N and parameter N_s^{-1} , the progeny variances read

$$\sigma_s^2 = \mu_s(1 - N_s^{-1}) . \tag{10}$$

Substituting (10) into (9) leads to

$$N_e^{(i)} = \frac{9N_1N_2}{2(2N_1 + N_2)} . \tag{11}$$

WRIGHT (1933), MALÉCOT (1951) and KIMURA (1963) used path coefficients, identity by descent and identity in state, respectively, to deduce (11). Observe that (11) cannot exceed $\frac{1}{2}N_1 + N_2$; equality occurs if and only if $N_1 = N_2 = \frac{1}{2}N$, in which case $N_e^{(i)} = 3N/4$. In the limits $N_1 \ll N_2$ and $N_1 \gg N_2$, the approximate values of (11) are $9N_1/2$ and $9N_2/4$, respectively. Since empirically the gametic variances exceed the binomial value (10) (CROW and MORTON 1955) and (9) is a decreasing function of each variance, it follows that the inbreeding effective population number is usually less than (11).

THE EXPECTED HOMOZYGOSITY

We suppose that every allele mutates at rate u to types not preexisting in the population. Evidently, then, we must multiply the right-hand side of each equation in the system (1) by $(1-u)^2$. In view of (4), we conclude immediately that the probabilities of identity in state converge to the solution of

$$\hat{f} = (1-u)^2 \hat{g} \quad (12a)$$

$$\hat{g} = (1-u)^2 \left\{ \frac{1}{2} \hat{g} + \frac{1}{2} [P_2 \frac{1}{2} (1+\hat{f}) + (1-P_2)\hat{l}] \right\} \quad (12b)$$

$$\hat{k} = (1-u)^2 [P_2 \frac{1}{2} (1+\hat{f}) + (1-P_2)\hat{l}] \quad (12c)$$

$$\begin{aligned} \hat{l} = (1-u)^2 \left\{ \frac{1}{4} [P_1 + (1-P_1)\hat{k}] + \frac{1}{4} [P_2 \frac{1}{2} (1+\hat{f}) \right. \\ \left. + (1-P_2)\hat{l}] + \frac{1}{2} \hat{g} \right\} \quad (12d) \end{aligned}$$

at the approximate ultimate rate

$$(1-u)^2 [1 - (2N_e^{(i)})^{-1}]^t \approx \exp\{-[2u + (2N_e^{(i)})^{-1}]t\} \quad (13)$$

We solve (12) for the probability that at equilibrium a female is homozygous:

$$\hat{f} = \frac{(1-u)^4 [2P_2 + P_1(1-P_2)(1-u)^2]}{[2 - (1-u)^2] \{4 - (1-P_2)(1-u)^2 [(1-P_1)(1-u)^2 + 1]\} - 2(1-u)^4} \quad (14)$$

Expanding (14) and recalling (5) informs us that for weak mutation and large population size

$$\hat{f} = \frac{1 + O(u, N^{-1}, Nu^2)}{1 + 4N_e^{(i)}u} \quad (15)$$

It is easy to see that \hat{g} , \hat{k} and \hat{l} are also given by (15).

Thus, provided we employ the appropriate effective population number (9), the expected equilibrium homozygosity has the same form as in the autosomal case, which was analyzed by MALÉCOT (1946, 1948) for identity by descent and sampling with replacement and by KIMURA and CROW (1964) and CROW and KIMURA (1970, pp. 332-327) for identity in state and an arbitrary progeny distribution. MALÉCOT (1951) also established (15) for an X-linked locus in the special case of identity by descent and binomial sampling.

A less detailed argument, which applies unmodified to an autosomal locus, will reveal the reason for the generality of the approximation

$$\hat{f} \approx (1 + 4N_e^{(i)}u)^{-1} \quad (16)$$

In terms of the probabilities of nonidentity, (12) has the form

$$\hat{\mathbf{V}} = [1 - (1-u)^2] \mathbf{1} + (1-u)^2 B \hat{\mathbf{V}},$$

in which $\mathbf{1}$ represents the vector of ones. Consequently,

$$\hat{\mathbf{V}} = [1 - (1-u)^2][I - (1-u)^2 B]^{-1} \mathbf{1}, \quad (17)$$

I being the identity matrix. Now, B deviates from a stochastic matrix by $O(P)$, where $P = P_1 + P_2$. Since $\mathbf{1}$ is the dominant eigenvector of this stochastic matrix, $\hat{\mathbf{V}}$ differs by $O(P)$ from the dominant eigenvector of B . Hence, we may approximate (17) for $P \rightarrow 0$ as

$$\hat{\mathbf{V}} \approx [1 - (1-u)^2][1 - (1-u)^2 \lambda_0]^{-1} \mathbf{1}, \quad (18)$$

Recalling the definition (4), we deduce for $u \ll 1$ from the first component of (18) that

$$\hat{h} \approx 4N_e^{(i)} u (1 + 4N_e^{(i)} u)^{-1},$$

which is equivalent to (16).

LITERATURE CITED

- CHIA, A. B. and E. POLLAK, 1974 The inbreeding effective number and the effective number of alleles in a population that varies in size. *Theoret. Pop. Biol.* **6**: 149-172.
- CROW, J. F., 1954 Breeding structure of populations. II. Effective population number. pp. 543-556. In: *Statistics and Mathematics in Biology*. Edited by O. KEMPTHORNE, T. A. BANCROFT, J. W. GOWEN and J. L. LUSH. Iowa State University Press, Ames, Iowa.
- CROW, J. F. and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- CROW, J. F. and N. E. MORTON, 1955 Measurement of gene frequency drift in small populations. *Evolution* **9**: 202-214.
- ETHIER, S. N. and T. NAGYLAKE, 1980 Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Prob.* **12**: 14-49.
- HARTL, D. L., 1971 Some aspects of natural selection in arrhenotokous populations. *Am. Zoologist* **11**: 309-325.
- KIMURA, M., 1963 A probability method for treating inbreeding systems, especially with linked genes. *Biometrics* **19**: 1-17.
- KIMURA, M. and J. F. CROW, 1963 The measurement of effective population number. *Evolution* **17**: 279-288. —, 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725-738.
- MALÉCOT, G., 1946 La consanguinité dans une population limitée. *Compt. Rend. Acad. Sci.* **222**: 841-843. —, 1948 *Les mathématiques de l'hérédité*. Masson, Paris. (Extended translation: *The Mathematics of Heredity*. Freeman, San Francisco, 1969.) —, 1951 Un traitement stochastique des problèmes linéaires (mutation, linkage, migration) en Génétique de Population. *Ann. Univ. Lyon, Sci., Sec. A.*, **14**: 79-117.
- MORAN, P. A. P. and G. A. WATTERSON, 1959 The genetic effects of family structure in natural populations. *Australian J. Biol. Sci.* **12**: 1-15.

- POLLAK, E., 1977 Effective population numbers and their interrelations. pp. 115-144. In: *Proceedings of the Washington State University Conference on Biomathematics and Biostatistics of May 1974*. Department of Pure and Applied Mathematics, Washington State University and Pi Mu Epsilon, Washington Alpha Chapter, Pullman, Washington.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97-159. —, 1933 Inbreeding and homozygosis. *Proc. Natl. Acad. Sci. U.S.* **19**: 411-420. —, 1938 Size of population and breeding structure in relation to evolution. *Science* **87**: 430-431. —, 1939 *Statistical Genetics in Relation to Evolution*. Actual. Sci. Ind. **802**: Exposés de biométrie et de statistique biologique. Edited by G. TEISSIER. Hermann, Paris. —, 1949 Genetics of populations. *Encyclopaedia Britannica*, 14th ed., **10**: 111-112. —, 1978 *Evolution and the Genetics of Populations*, Vol. IV. The University of Chicago Press, Chicago.

Corresponding editor: W. J. EWENS