

EFFECT OF SELECTION AND LINKAGE ON INBREEDING*

By T. P. BOGYO† and SHIRLEY W. TING†

[Manuscript received March 17, 1967]

Summary

Gill and Clemmer performed an inbreeding experiment by using Monte Carlo simulation techniques and reported that both selection and linkage significantly increased the rate of inbreeding in a population of a given size. The results could not be explained by the theoretical explanation that linkage should not affect inbreeding coefficient in the absence of selection. A programme was written to repeat Gill's and Clemmer's experiment.

The results of our experiment are different from those reported by Gill and Clemmer. It is assumed that there was a randomization error in their programme since, by introducing a very simple randomization error, we could show a significant increase in the inbreeding coefficient even in the absence of selection. When applying correct randomization procedures we obtained results very much similar to what one would expect from theory. Linkage in the absence of selection does not produce any effect on the inbreeding coefficient but does affect the variance of the estimates (just as one would expect from the results of Schnell). Truncation selection has an increasing effect on the inbreeding coefficient, but does not affect the variance of the estimate. With both linkage and truncation selection active an interaction will be evident in the more advanced generations of inbreeding.

I. INTRODUCTION

Gill and Clemmer (1966) performed an experiment using Monte Carlo simulation techniques and reported that both selection and linkage significantly increased the rate of inbreeding in populations of given size. They found the effect of linkage to be a significant factor in inflating the average inbreeding coefficient of individuals of the progeny population. This astonishing result could not be explained by theoretical justification; in fact all theoretical considerations demand that all computed inbreeding coefficients should have the same expectation, irrespective of linkage. Since there was a possibility that Gill and Clemmer could have been in error by some systematic bias introduced in the simulation programme, it was decided to repeat the experiment with an entirely fresh programme written specifically for the purpose. The following results will show that linkage *does not* have any effect on increasing the inbreeding coefficient and will also show how a very small systematic bias in the simulation programme can introduce an error similar to that obtained by Gill and Clemmer.

* Scientific paper No. 2952, Washington State University, College of Agriculture Research Centre, Pullman. Project Nos. 1775 and 1884. Supported in part by National Science Foundation Grant No. GB-5472 and Washington State Initiative No. 171.

† Washington State University, Pullman, Washington 99163.

II. METHODS

Two sets of experiments were carried out: one with a parent population size (N) of four (two males and two females) and another one with a parent population size of eight (four males and four females). Each experiment was a $3 \times 3 \times 2$ factorial with the following levels: three levels of linkage (0, 0.60, and 0.99) corresponding to the recombination probabilities of 0.5, 0.2, and 0.005 respectively; three levels of parent-offspring population size contrasts (1 : 1, 1 : 2, and 1 : 4); and two modes of selection—truncation selection and random selection. The level of linkage was the same for all the loci on every chromosome. The difference between the modes of selection was simulated by either reordering the offspring population before selecting parents or by selecting without reordering.

Each individual in the population had five pairs of chromosomes with five loci on each one. The initial gene frequency at every locus of the original parent population was generated to be 0.5, although no effort was made to force the initial population to be in Hardy-Weinberg equilibrium. Gamete production was achieved by using a binary mask and since this masking procedure is a potential source of bias, some clarification about the procedure seems to be in order.

The method of producing a single gamete involves testing a random number against the vector of test frequencies. Since the possible number of different loci and also the possible number of crossover types is 2^n , where n is the number of loci per chromosome, the length of the vector is 2^n , in this experiment 32. The crossover mask to be used is chosen by testing a random number in the range 0–1 against each of the precalculated test frequencies in turn until the requirement that the random number have a value less than the test frequency is satisfied. The index of the test frequency will then select the binary mask which is used to operate on the parent genotype. The operation consists of calculating the logical product of the mask and one of the chromosomes and logically adding this to the logical product of the complement of this mask and the pair of the chromosome chosen in the first place.

In order to evaluate the estimated inbreeding coefficient (F) all parental genes were identified uniquely. Genes were represented by a binary bit in the computer, thus plus genes were identified as 1's and minus genes as 0's. Each computer word (36 bits long) represented a chromosome. The rightmost five bits were used to represent the five genes in question, the sixth bit was always empty and set to zero. The next 15 bits were used to identify individual genes in sets of three (since the largest parental population was eight), the leftmost 15 bits were used to distinguish between the homologous chromosome pairs, thus making the identification completely unique. The crossover masks were designed so that when a crossover occurred the same crossover also carried the identification bits simultaneously to the gamete word. In evaluating the inbreeding coefficient the total number of identical gene pairs were counted (again by using masks) in the progeny population and F was estimated as the proportion of these pairs out of the total number of possible pairs (25 times the progeny population size).

A completely additive model (no dominance, no epistasis) was used and the phenotypic value was made equal to the sum of the plus genes per individual. The

value of a single individual thus varies between 0 and 50. For truncation selection the individuals were ranked according to their phenotypic value (which, since no environmental effect was simulated, equals the genotypic value); for random selection no ranking was performed. The first p proportion (where p is the selection pressure) of the offspring population was chosen to be parents of the next generation, with the first $N/2$ being males and the second $N/2$ being females.

In order to investigate the source of a systematic bias in the sampling procedures two different versions of the simulating programme were run. In the first version the crossover mask was always applied to chromosome I of the individual and the complement of the mask applied to chromosome II. In a second version this procedure was changed and the original mask was applied at random to either the first or to the second chromosome.

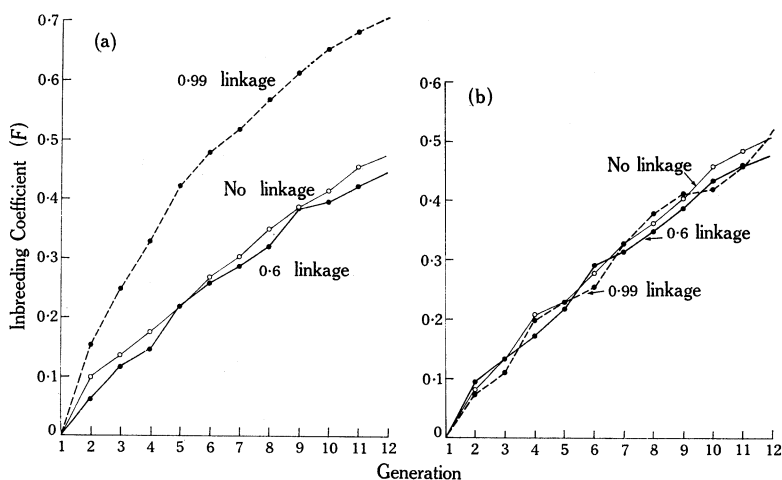


Fig. 1.—Mean inbreeding by levels of linkage (no selection) for populations of eight parents with incorrect (a) and correct (b) randomization.

III. RESULTS

While the first version of the programme gave both F values and F variances much lower than those obtained by Gill and Clemmer, the results still showed a significant effect of linkage on inbreeding, even in the absence of selection. Possible sources of error were then considered and one of them, namely a procedure which could effect the random choice of gametes, was considered as a potential source of error, since a bias in choosing the gametes could reduce the effective population size considerably. A different randomization procedure, the second version, was then introduced and this changed the results dramatically. All effects of linkage in the absence of selection disappeared. Figures 1(a) and 1(b) contrast the differences between the results of the two simulating procedures. All future references apply to results obtained by the "correct" simulating procedure. Since results for parental populations of size four were similar to those with size eight, results for only the latter will be shown.

TABLE 1
ANALYSIS OF VARIANCE OF INBREEDING COEFFICIENTS BY GENERATIONS FOR A PARENT POPULATION SIZE OF EIGHT

Source of Variation	Degrees of Freedom	Mean Squares for Generations:										
		2	3	4	5	6	7	8	9	10	11	12
Mode of selection (<i>M</i>)	1	0.0120*	0.0161**	0.0808**	0.1239**	0.1796**	0.3263**	0.4163**	0.6197**	0.7235**	0.9425**	1.1447**
Linkage (<i>L</i>)	2	0.0016	0.0001	0.0018	0.0081	0.0042	0.0174*	0.0538**	0.0668**	0.0541**	0.0855**	0.1383**
Parent-offspring population size contrast (<i>S</i>)	2	0.0086**	0.0229**	0.0308**	0.0723**	0.0837**	0.0874**	0.1467**	0.1739**	0.2068**	0.2652**	0.2322**
Replication (<i>R</i>)	19	0.0022	0.0025	0.0026	0.0029	0.0027	0.0017	0.0038	0.0029	0.0023	0.0029	0.0026
Interaction												
<i>M</i> × <i>L</i>	2	0.0006	0.0003	0.0045	0.0051	0.0012	0.0134*	0.0219*	0.0171*	0.0089	0.0206	0.0474**
<i>M</i> × <i>S</i>	2	0.0021	0.0336**	0.0136*	0.0612**	0.0771**	0.1140**	0.1402**	0.1651**	0.1990**	0.2454**	0.2717**
<i>L</i> × <i>S</i>	4	0.0039	0.0040	0.0061	0.0049	0.0075	0.0097*	0.0132*	0.0080	0.0220**	0.0115*	0.0098
<i>M</i> × <i>L</i> × <i>S</i>	4	0.0013	0.0066*	0.0032	0.0058	0.0185**	0.0097*	0.0151*	0.0164*	0.0181**	0.0340**	0.0294**
Error	323	0.0018	0.0022	0.0031	0.0035	0.0038	0.0039	0.0050	0.0049	0.0047	0.0047	0.0047

P* < 0.05.*P* < 0.01.

Table 1 is the combined analysis of variance of simulated F values. It shows very clearly that, in the second generation of breeding, linkage (over all levels of population size and mode of selection) is completely ineffective in changing the level of inbreeding. Mode of selection (truncation *v.* random), on the other hand, is a most powerful factor in influencing the inbreeding coefficient and remains such over all generations (Fig. 2).

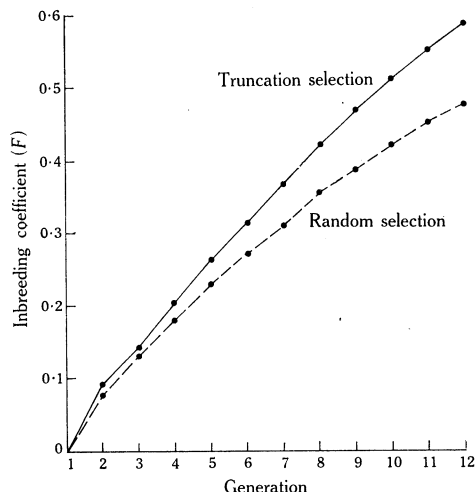


Fig. 2.—Mean inbreeding by modes of selection over all levels of linkage and parent-offspring population size contrasts for populations of eight parents.

Parent-offspring population size contrast (over all levels of linkage and modes of selection), as expected, is the other factor which has significant effect on the inbreeding coefficient, starting from the second generation and continuing over all generations of breeding [Fig. 3(a)]. Because of the highly significant interaction between modes of selection and parent-offspring population size contrasts, Figure 3(b) gives the curve for the contrasts for truncation selection over all levels of linkage, and Figure 3(c) the same curves with random selection. All lines follow the theoretical curve for $N_e = N = 8$, as expected.

It has already been established from Figure 1(b) that, in the absence of selection, linkage has no effect on influencing the inbreeding coefficient, and the curve over generations follows the theoretical curve exactly. However, as the analysis of variance table shows, averaged over all modes and levels of selection a significant linkage effect appears in the seventh generation of breeding and this effect becomes more and more evident generation after generation beyond the seventh [Fig. 3(d)]. This difference is almost entirely due to the highest level of linkage (99%), the 60% level following the "no linkage" line almost exactly. Since this line is an average line over all levels of selection, some of which have been shown (and are known) to increase the inbreeding coefficient, one would expect this line to be somewhat above that of the theoretical one. The 99% line is not a simple function of the effective parent

population size

$$F_t = (1/2N)(1 + F_{t-2}) + [1 - (1/N)]F_{t-1},$$

but seems to be influenced also by the increased gene frequencies caused by selection.

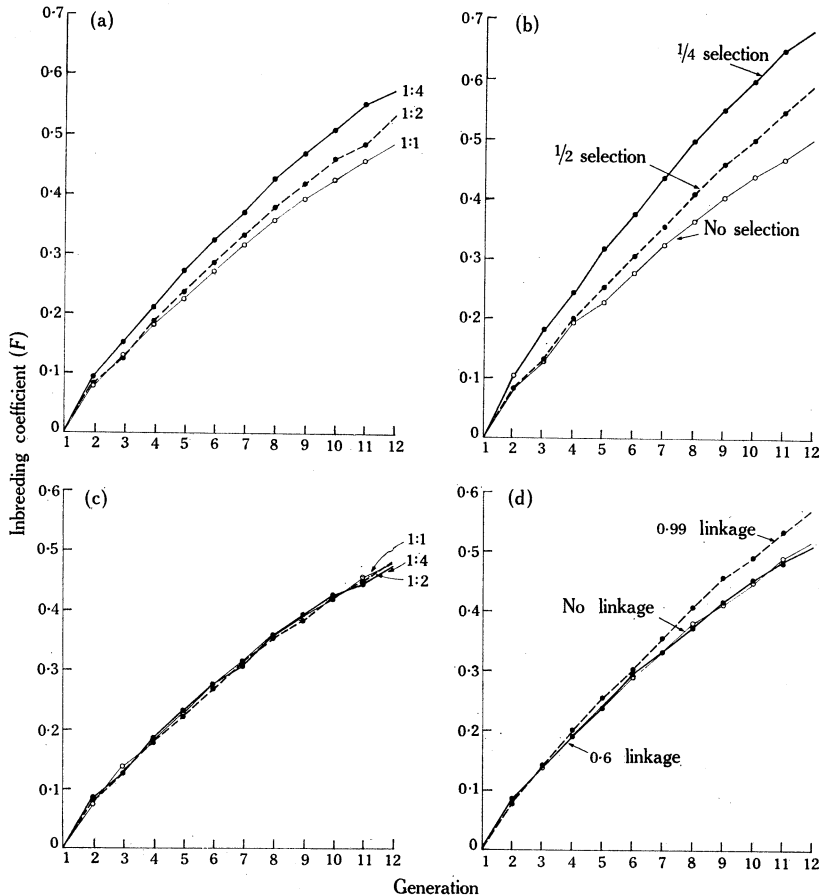


Fig. 3.—(a)–(c) Mean inbreeding by parent–offspring population size contrasts over all levels of linkage for populations of eight parents for (a) all modes of selection, (b) truncation selection (correct randomization), and (c) random selection. (d) Mean inbreeding by levels of linkage over all modes and levels of selection for populations of eight parents.

In the case of very tight linkage this seems to favour entire chromosomes rather than genes and in effect reduces the effective population size. Further evidence of this is shown by the diagram of the variances of the inbreeding coefficients over generations [Fig. 4(a)]. The variances of the inbreeding coefficient estimates in the absence of selection show a slow but general increase in the variances over generations with populations of size $N = 8$ [Fig. 4(b)] and an increase followed by a decrease for parent

populations of size $N = 4$ [Fig. 4(c)]. This confirms that the expected values of these variances are functions of $F(1-F)$. In both population sizes tight linkage is expected to increase the variances as pointed out by Schnell (1963). The variances of the inbreeding coefficient estimates are not affected by selection, irrespective of whether they are averaged over all levels of linkage [Fig. 4(d)] or not [Fig. 4(e)].

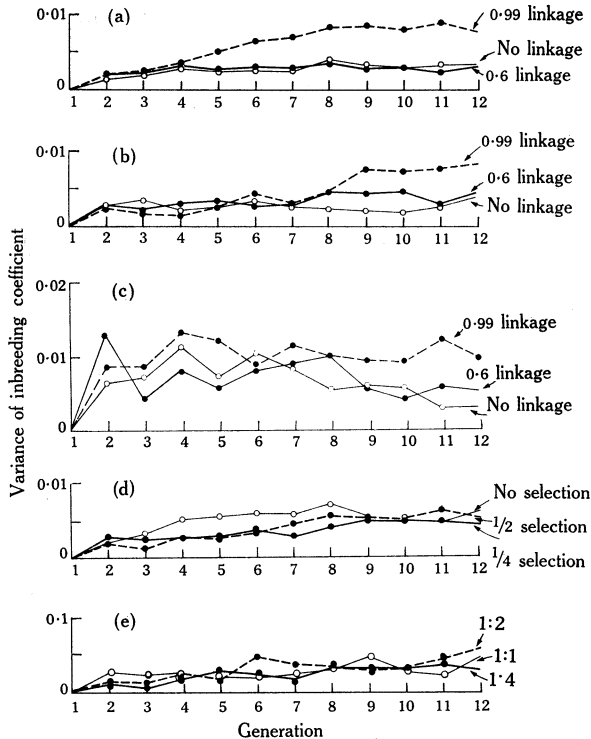


Fig. 4.—(a) and (b) Variance of inbreeding coefficient by levels of linkage for populations of eight parents (correct randomization) over all levels of selection (a) and with no selection (b). (c) As for (b) but for populations of four parents. (d) and (e) Variance of inbreeding coefficient for populations of eight parents (correct randomization) by levels of selection over all levels of linkage (d) and by levels of random selection with no linkage (e).

IV. DISCUSSION

The primary concern of this study was the effect of linkage on the inbreeding coefficient in the absence of selection. The theoretical expected F in small populations with no selection has been given by Wright (1931) as:

$$F_n = (1/2N)(1 + F_{n-2}) + [1 - (1/N)]F_{n-1}. \quad (1)$$

Gill and Clemmer (1966) in their simulation experiment compute the inbreeding coefficient as an average value over all loci and all individuals in a particular generation. In the absence of selection, irrespective of linkage, the expected value of F

remains as given by equation (1). This can be shown as follows: Let $F_{i(0)}$ be the expectation of the computed inbreeding coefficient for no recombination at the i th locus and $F_{i(0.2)}, F_{i(0.5)}$ be the corresponding expectations for 0.2 and 0.5 values of recombination. Since normal segregation and random union of gametes is assumed, $F_{i(0)}$ must be equal to $F_{i(0.2)}$ and $F_{i(0.5)}$ because what happens to other loci is completely ignored. Thus, any estimate $\hat{F}_{i(0)}, \hat{F}_{i(0.2)}, \hat{F}_{i(0.5)}$ by taking an identity by descent over individuals in any particular generation must have the same expected value, F_i . In the present study the counting procedure was to calculate first the average inbreeding coefficient over all loci ($i = 1, \dots, n$) for the k th replicate ($k = 1, \dots, r$) for any particular level of linkage (q) as

$$\bar{F}_{qk} = \sum_i \hat{F}_{ik} / n,$$

and, after all replications have been completed, to average over all replications

$$\bar{F}_q = \sum_k \bar{F}_{qk} / r.$$

Averaging $\hat{F}_{ik}$ over replications first and then over all loci will furnish *identical* results.

Selection, of course, will increase F because it causes the effective number of parents to be less than the census number. An approximate formula, of the ratio of census number to effective number, based on heritability, was given by Robertson (1961) as

$$N/N_e = 1 + 2\bar{i}^2 h^2 [1 - h^2 \bar{i}(\bar{i} - x)], \quad (2)$$

where x is the abscissa of the unit normal curve at the point of truncation, and $\bar{i}$ is the standardized selection differential.

The effect of linkage on the variance of F was discussed by Schnell (1961, 1963). He showed that the distribution of m , the number of loci homozygous by descent after a period of inbreeding is related to the inbreeding function and that the variance of $\bar{m}$, the mean number of loci, homozygous by descent, consists of two parts, one which is equivalent to $nF(1 - F)$ and another which gives the contribution due to linkage. Using Schnell's parameters, F_{ij} , the probability of two given genes being simultaneously identical by descent (which in the absence of linkage should be equal to F^2) it can be shown that the variance of estimates of F , $\hat{F}$, is affected by linkage, and inbreeding will increase this variance. Let F_{ij} be the frequency with which genes are identical at the i th locus and not identical at the j th, $F_{i'j}$ be the frequency with which the genes are identical at the j th locus but not at the i th and $F_{i'j'}$ the frequency with which the genes are *not* identical at both loci. With complete linkage ($q = 0$) with each gene being identified uniquely in the base population, $F_{i'j(0)} = F_{ij'(0)} = 0$ since a recombination is required for these frequencies to be different from zero. Moreover, for $q = 0$, $\hat{F}_{ij(0)} = \hat{F}_{i(0)} = \hat{F}_{j(0)}$ for every estimate. Consequently

$$\sigma^2_{0.5(\hat{F}_{i(0)} + \hat{F}_{j(0)})} = \sigma^2_{(\hat{F}_{ij(0)})} = \sigma^2_{(\hat{F}_{i(0)})} = \sigma^2_{(\hat{F}_{j(0)})} = \sigma^2_{(\hat{F}_{\cdot(0)})}.$$

With free recombination ($q = 0.5$), however,

$$\hat{F}_{ij(0.5)} = F_{i(0.5)}F_{j(0.5)} = F^2_{(0.5)},$$

and

$$\sigma^2_{0.5(\hat{F}_{i(0.5)} + \hat{F}_{j(0.5)})} = \frac{1}{4}[\sigma^2(\hat{F}_{i(0.5)}) + \sigma^2(\hat{F}_{j(0.5)})] = \frac{1}{2}\sigma^2(\hat{F}_{(0.5)}),$$

but for any particular generation $\sigma^2(\hat{F}_{(0)}) = \sigma^2(\hat{F}_{(0.5)}) = \sigma^2(\hat{F})$. In the case of complete linkage the loci estimates are perfectly correlated, the general result for m completely linked loci being

$$\sigma^2(\sum_i \hat{F}_{i(0)})/m = \sigma^2(\hat{F}),$$

while for m independent loci

$$\sigma^2(\sum_i \hat{F}_{i(0.5)})/m = \sigma^2(\hat{F})/m.$$

To avoid further misunderstanding it should again be emphasized that a clear distinction must be made between F (the probability of a single locus being identical by descent) and F_{ij} (the joint probability of two or more genes being identical by descent). The variance of F for a particular locus among unrelated individuals or among unrelated groups of individuals is a function of $F(1 - F)$ alone and this variance is not affected by linkage. Linkage will, however, affect F_{ij} and this will lead to a correlation of the actual F values among loci. It will reduce the variation among F values among loci of the same individuals or of averages of groups such as individual computer runs. Since the F 's among loci are correlated within computer runs, averages of them will have greater variances among computer runs. These results follow directly from the results of Schnell (1963). While it may be difficult to quantify, it seems logical that selection will augment this increase in the variances because it will increase F .

V. ACKNOWLEDGMENTS

The authors are indebted to Dr. C. Clark Cockerham for his helpful criticism during the preparation of this manuscript and to Dr. A. S. Fraser for many valuable discussions.

VI. REFERENCES

- GILL, J. L., and CLEMMER, B. A. (1966).—Effects of selection and linkage on degree of inbreeding. *Aust. J. biol. Sci.* **19**, 307–17.
- GRIFFING, B. (1960).—Accommodation of linkage in mass selection theory. *Aust. J. biol. Sci.* **13**, 501–26.
- ROBERTSON, A. (1961).—Inbreeding in artificial selection programmes. *Genet. Res., Camb.* **2**, 189–94.
- SCHNELL, F. W. (1961).—Some general formulations of linkage effects in inbreeding. *Genetics, Princeton* **46**, 947–57.
- SCHNELL, F. W. (1963).—The covariance between relatives in the presence of linkage. In “Statistical Genetics and Plant Breeding”. pp. 468–83. (Publ. No. 982, Nat. Acad. Sci.—Nat. Res. Coun.)
- WRIGHT, S. (1931).—Evolution in Mendelian populations. *Genetics, Princeton* **16**, 97–159.

