

# THE EFFECT OF LINKAGE ON THE MEAN VALUE OF INBREDS DERIVED FROM A RANDOM MATING POPULATION<sup>1</sup>

I. M. R. VAN AARDE

*Statistical Laboratory, Iowa State University, Ames Iowa 50010*

Manuscript received February 1, 1973

Revised copy received June 3, 1974

## ABSTRACT

An expression is derived which accounts for the effect of linkage on the mean value of diploid inbreds. The original population is taken to be in Hardy-Weinberg equilibrium. It is shown that linkage will accelerate inbreeding depression. The precise nature of the acceleration is worked out for some special cases.

THE purpose of this paper is to exhibit an expression for the mean genotypic value of some trait in a population that has been derived (without selection) from a Hardy-Weinberg population of diploid ancestors where there is arbitrary linkage epistasis. It is assumed that there are no position effects.

In the case of no linkage, KEMPTHORNE (1957) has shown that the mean of the inbred population deviates from the mean of the ancestors by a quantity which he expresses in terms of Wright's  $F$  as

$$FD_1 + F^2D_2 + F^3D_3 + \dots + F^nD_n.$$

Here  $n$  is the number of loci, and  $D_k$  involves only the "dominance x dominance x . . . x dominance" deviations from  $k$  loci, as defined for the ancestral population. The coefficient of  $D_k$  arises from the product rule for the probability of  $k$  independent events as  $F \times F \times F \times \dots \times F$  from  $k$  factors.

With linkage, however, the probability that the genes of an inbred individual are identical by descent at both of two tightly linked loci is more nearly  $F$  than  $F \times F$ . Hence, their joint contribution to inbreeding depression will be similar to  $D_1$  rather than to  $D_2$ . Similar reasoning applies to a block of  $k$  tightly linked loci. The aim is to show precisely what the relationships are. This is accomplished by deriving a result which amounts to replacing the  $k$ -th power of the one-locus inbreeding coefficient in KEMPTHORNE'S expression by the  $k$ -locus generalized coefficient to account for linkage. As such, the result is almost self-evident. Therefore, KEMPTHORNE'S development and notation is used as far as possible.

<sup>1</sup> Journal Paper No. J-7493 of the Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 1669. The research reported here was supported partially by the National Science Foundation, Grant No. 19218 and the National Institutes of Health, Grant No. 13827.

<sup>2</sup> Present address: Faculty of Agriculture, University of Stellenbosch, Republic of South Africa.

*The Mean Of An Inbred Population*

Let random individuals be drawn from the ancestral population and let matings be made so as to derive (without selection of any sort) an inbred individual,  $X$ . A given mating procedure specifies an infinite population of inbreds from which  $X$  is a random individual. We shall now derive a general expression for the mean of such a population of inbreds.

The derivation is facilitated by working with the generalized inbreeding coefficient of  $X$  with respect to a given set of loci: This is the probability that at every locus in the set, the genes that  $X$  receives from its parents ( $X_0$  and  $X_1$ ) are identical by descent. For the inbreeding coefficients of  $X$  with respect to the sets of loci

$$\{i\}, \{i,j\}, \{i,j,k\}, \dots,$$

we make the notations

$$F_i, F_{ij}, F_{ijk}, \dots,$$

respectively.

Consider the two genes of  $X$  at locus  $i$ . We make the notations

$$A_{x_0^i}^i = (\text{the gene that } X \text{ receives from } X_0), \text{ and}$$

$$A_{x_1^i}^i = (\text{the gene that } X \text{ receives from } X_1).$$

We extend the notation by writing

$$(A_{x_0^i}^i, A_{x_1^i}^i) \equiv (A_{z_0^i}^i, A_{z_1^i}^i)$$

when the genes are copies of *different* ancestral genes (indexed by  $z_0^i$  and  $z_1^i$ ), and

$$(A_{x_0^i}^i, A_{x_1^i}^i) \equiv (A_{z^i}^i, A_{z^i}^i)$$

when the genes are identical by descent, both being copies of a *single* ancestral gene (indexed by  $z^i$ ). The ancestral genes are random; that is to say, each one may be identical in state to any one of an arbitrary number of allelomorphic forms,

$$A_1^i, A_2^i, A_3^i, \dots,$$

which occur at locus  $i$  with arbitrary frequencies,

$$p_1^i, p_2^i, p_3^i, \dots,$$

respectively. Hence we obtain  $n$  pairs of random variables,

$$(z_0^1, z_1^1), (z_0^2, z_1^2), (z_0^3, z_1^3), \dots, (z_0^n, z_1^n),$$

such that their joint distribution is characterized by the following four properties:

- i) Members of different pairs are independent because the ancestral population is in Hardy-Weinberg equilibrium.
- ii) Members of the same pair are (for the same reason) independent when they index different ancestral genes.
- iii) Members of the same pair are aliases of a single random variable ( $z^i$  for pair  $i$ ) when they index the same ancestral gene.

iv) Members of the triplet  $(z_0^i; z_1^i; z^i)$  always have the same marginal distribution, given by the probabilities  $P(z^i = j) = p_j^i$  for all  $(i, j)$ . The genes of  $X$  at locus  $i$  can now be characterized by the equations

$$\begin{aligned} A_{x_0^i}^i &= (1 - \delta^i) A_{z_1^i}^i + \delta^i A_{z^i}^i, \text{ and} \\ A_{x_1^i}^i &= (1 - \delta^i) A_{z_0^i}^i + \delta^i A_{z^i}^i, \end{aligned}$$

where we introduce  $n$  random Kronecker deltas,

$$\delta^1, \delta^2, \delta^3, \dots, \delta^n,$$

such that

$$\begin{aligned} P(\delta^i = 1) &= F_i, \\ P(\delta^i = 1, \delta^j = 1) &= F_{ij}, \\ P(\delta^i = 1, \delta^j = 1, \delta^k = 1) &= F_{ijk}, \end{aligned}$$

and so on.

We now express the genotypic value of  $X$  in terms of the formal product

$$\prod_{i=1}^n (1 + \alpha_{x_0^i}^i) (1 + \alpha_{x_1^i}^i) \tag{1}$$

which has precisely the same meaning as the corresponding expression used by KEMPTHORNE (1957); we have merely adjoined indices to identify the genomes that  $X$  receives from  $X_0$  and  $X_1$ , respectively. Precisely as in KEMPTHORNE'S development, the expression must be expanded into a sum of  $2^{2n}$  terms. The term unity represents the mean of the random bred population of ancestors,  $\mu_R$ . The remaining  $2^{2n} - 1$  terms represent deviations. For instance,

$$\alpha_{x_1^i}^i \alpha_{x_0^j}^j \alpha_{x_0^k}^k \alpha_{x_1^k}^k$$

represents the "additive  $\times$  additive  $\times$  dominance" deviation associated with

- gene  $A_{x_1^i}^i$  at locus  $i$ ,
- gene  $A_{x_0^j}^j$  at locus  $j$ , and
- genes  $A_{x_0^k}^k$  and  $A_{x_1^k}^k$  at locus  $k$ ,

which is  $(a^i a^j d^k)_{x_1^i x_0^j x_0^k x_1^k}$  in KEMPTHORNE'S notation.

The  $i$ -th factor in (1) must be

$$\begin{aligned} (1 + \alpha_{x_0^i}^i) (1 + \alpha_{x_1^i}^i) &\text{ when } \delta^i = 0, \text{ and} \\ (1 + \alpha_{z^i}^i) (1 + \alpha_{z^i}^i) &\text{ when } \delta^i = 1. \end{aligned}$$

Therefore, we express the factor as

$$[1 + (1 - \delta^i) \alpha_{z_0^i}^i + \delta^i \alpha_{z_1^i}^i] [1 + (1 - \delta^i) \alpha_{z_1^i}^i + \delta^i \alpha_{z_0^i}^i].$$

Note that when we expand the product, ambiguous terms involving both  $z_0^i$  and  $z^i$  or both  $z_1^i$  and  $z^i$  vanish, because  $(1 - \delta^i)\delta^i$  and  $\delta^i(1 - \delta^i)$  vanish. More generally, we replace (1) with the formal product

$$\prod_{i=1}^n [1 + (1 - \delta^i) \alpha_{z_0^i}^i + \delta^i \alpha_{z_1^i}^i] [1 + (1 - \delta^i) \alpha_{z_1^i}^i + \delta^i \alpha_{z_0^i}^i]. \tag{2}$$

which must be expanded into a sum of  $7^n$  non-vanishing terms. The term unity

represents  $\mu_R$ . The remaining terms involve  $\alpha$ 's or mixed products of  $\alpha$ 's which represent deviations.

The random variables  $z_0^i, z_1^i, z^i$ , for  $i = 1, 2, 3, \dots, n$ , are statistically independent of the Kronecker deltas, because (with no selection) the state of an ancestral gene has no bearing on contingencies which may cause copies of that gene to be transmitted to descendants. Hence, to obtain the mean of the inbred population,  $\mu_I$ , we expand (2), identify the representations of  $\mu_R$  and the deviations, and take the expectation in two stages as follows: First we take the expectation only with respect to

$$\{z_0^i; z_1^i, z^i \mid i = 1, 2, 3, \dots, n\}.$$

Properties (i), (ii), (iii) and (iv) show that all terms involving one or more of the variables

$$\{z_0^i, z_1^i \mid i = 1, 2, 3, \dots, n\}$$

will vanish, leaving only  $\mu_R$  and terms involving quantities like

$$\begin{aligned} d_i &= E [(d^i)_{z^i z^i}], \\ d_{ij} &= E [(d^i d^j)_{z^i z^i z^j z^j}], \\ d_{ijk} &= E [(d^i d^j d^k)_{z^i z^i z^j z^j z^k z^k}], \end{aligned}$$

and so on. The second stage consists of taking the expectation with respect to

$$\{\delta^i \mid i = 1, 2, 3, \dots, n\}.$$

Here we simply have to note that

$$(\delta^i)^2 = \delta^i \text{ for all } i,$$

and

$$E(\delta^i \delta^j \delta^k \dots) = F_{ijk\dots} \text{ for all } i, j, k, \dots$$

Then we obtain the general result

$$\mu_I = \mu_R + \sum_i F_i d_i + \sum_{i < j} F_{ij} d_{ij} + \sum_{i < j < k} F_{ijk} d_{ijk} + \dots$$

The coefficients  $F_i$  take the same value, Wright's  $F$ , for all loci. With no linkage,

$$F_{ijk\dots} = F \times F \times F \times \dots,$$

so that the result reduces to the form given by KEMPTHORNE (1957) who makes the notations

$$D_1 = (\sum_i d_i), D_2 = (\sum_{i < j} d_{ij}), D_3 = (\sum_{i < j < k} d_{ijk}), \dots$$

### The Effect Of Linkage In General

Linkage tends to accelerate the effects which progressive inbreeding may have on the population mean, because

$$F^{ij} \geq F^2, F_{ijk} \geq F^3, F_{ijkl} \geq F^4, \dots$$

In particular, if loci 1 and 2 are very tightly linked (almost aliases)

$$\begin{aligned} F_{12} &\simeq F, \\ F_{12i} &\simeq F_{1j} \text{ and } F_{2i} \text{ for } i = 1, 2, 3, \dots, n, \\ F_{12ij} &\simeq F_{1ij} \text{ and } F_{2ij} \text{ for } i, j = 1, 2, 3, \dots, n, (i \neq j), \end{aligned}$$

and so on. Consequently,

$$(d_1 + d_2 + d_{12})$$

will approximate the contribution of a single locus to  $\mu_I$ ,

$$(d_{1i} + d_{2i} + d_{12i}) \text{ for } i = 1, 2, 3, \dots, n,$$

will approximate the contributions of pairs of loci to  $\mu_I$ ,

$$(d_{1ij} + d_{2ij} + d_{12ij}) \text{ for } i, j = 1, 2, 3, \dots, n, \quad (i \neq j),$$

will approximate the contributions of triplets of loci to  $\mu_I$ , and so on.

If loci  $1, 2, 3, \dots, k$ , are a tightly linked block of loci, the quantity

$$(d_1 + d_2 + d_{12} + d_3 + d_{13} + d_{23} + d_{123} + \dots + d_{123\dots k})$$

will approximate the contribution of a single locus to  $\mu_I$ .

### *The Effect Of Linkage In Special Cases*

The generalized inbreeding coefficients,  $F_i, F_{ij}, F_{ijk}, \dots$ , were introduced by SCHNELL (1961). In any given instance their values will depend on the mating procedure and the linkage relationships. For instance, if  $X$  is derived by one generation of selfing, SCHNELL obtains

$$\begin{aligned} F_i &= (1/2), \\ F_{ij} &= (1/2)^2 (1 + \lambda_{ij}^2), \text{ and} \\ F_{ijk} &= (1/2)^3 (1 + \lambda_{ij}^2 + \lambda_{ik}^2 + \lambda_{ijk}^2), \end{aligned}$$

wherein each  $\lambda$  is a linkage parameter defined as

$$1 - 2 \quad (\text{the corresponding recombination rate}).$$

Thus  $0 \leq \lambda_{ij} \leq 1$ , where  $\lambda_{ij}$  vanishes if loci  $i$  and  $j$  are unlinked and  $\lambda_{ij}$  approaches unity if loci  $i$  and  $j$  are tightly linked. Hence, for one generation of selfing,

$$\mu_I = \mu_R + (1/2) \sum_i d_i + (1/2)^2 \sum_{i < j} (1 + \lambda_{ij}^2) d_{ij} + \dots$$

Similarly, for one generation of parent-offspring mating,

$$\mu_I = \mu_R + (1/2)^2 \sum_i d_i + (1/2)^4 \sum_{i < j} (1 + \lambda_{ij}^2) (1 + \lambda_{ij}) d_{ij} + \dots,$$

and for one generation of half-sib mating,

$$\mu_I = \mu_R + (1/2)^3 \sum_i d_i + (1/2)^6 \sum_{i < j} (1 + \lambda_{ij}^2) (1 + \lambda_{ij})^2 d_{ij} + \dots$$

The coefficients of  $d_i$  decrease geometrically over the three mating procedures, but those of  $d_{ij}$  decrease less rapidly because  $(1 + \lambda_{ij}) \geq (1 + \lambda_{ij}^2)$ .

We note again that when  $\lambda_{12}$  approaches unity, the terms in  $d_1, d_2$ , and  $d_{12}$ , reduce to  $F (d_1 + d_2 + d_{12})$  which corresponds to the contribution of a single locus.

### LITERATURE CITED

- KEMPTHORNE, O., 1957 *An Introduction to Genetic Statistics*. John Wiley and Sons, Inc., New York.
- SCHNELL, F. W., 1961 Some general formulations of linkage effects in inbreeding. *Genetics* **46**: 947-957.

Corresponding editor: T. PROUT