ACCOMMODATION OF GENE–CHROMOSOME CONFIGURATION EFFECTS IN QUANTITATIVE INHERITANCE AND SELECTION THEORY

By B. GRIFFING*

[*Manuscript received January* 16, 1961]

Summary

Gene–chromosome configuration effects may be generated in at least two different ways. The first results from the position-effect phenomenon, and the second, which is manifest if the individual is evaluated on the basis of its inbred progeny, is due to the restriction of independent segregation because of linkage. The present study is an attempt to generalize the gene model used in quantitative inheritance and selection theory so that it may accommodate these effects.

Configuration effects are defined and their relationships to the effects in the conventional model are examined for a random-mating population in equilibrium. Then, the expectations of various covariances among relatives are developed for the complete model which includes the configuration effects. Finally, the importance of this extension is discussed, primarily from the point of view of artificial selection.

I. INTRODUCTION

It is clear that more than one gene–chromosome arrangement is possible for diploid organisms heterozygous for the same set of genes at two or more linked loci. Thus, for two linked loci, the two possible genotypes are $(A_1^1A_1^2)$ $(A_2^1A_2^2)$ and $(A_1^1A_2^2)$ $(A_2^1A_1^2)$, where A_i^j is the ith allele at the jth locus, and the gene content within each set of parentheses indicates the association of genes within each homologous chromosome. More generally, if there are n such loci, the number of different genotypes for the given set of genes is 2^{n-1}.

From the standpoint of quantitative inheritance and selection theory, there are at least two ways in which the gene–chromosome arrangement may influence the evaluation of a genotype.

First, the physical configuration of the genes may induce position effects. The position-effect phenomenon was first discovered in *Drosophila melanogaster* by Sturtevant (1925), and more recently it has been found in a wide range of organisms.

Second, if the various gene arrangements are evaluated on the basis of their inbred progeny, it is obvious that different configurations may yield different evaluations for the same set of genes. For example, with the simplest situation involving only two loci, the expected selfed progeny arrays are different for the genotypes $(A_1^1A_1^2)$ $(A_2^1A_2^2)$ and $(A_1^1A_2^2)$ $(A_2^1A_1^2)$ if the recombination value is less than one-half. Hence, the expected means of these progeny arrays may be different if epistasis occurs.

* Division of Plant Industry, C.S.I.R.O., Canberra.

In quantitative inheritance theory no attempt has been made to accommodate effects due solely to the gene–chromosome arrangement. Hence, the objective of this study is to generalize the gene model used in quantitative inheritance so that it may do so. This requires both a genotypic representation which permits all possible gene–chromosome arrangements to be distinguishable, and a definition of configuration effects which may be incorporated into the gene model.

In the next section, an appropriate genotypic representation is developed; configuration effects are defined and their relationships to the effects in the conventional model are examined for a random-mating population in equilibrium. Then, the expectations of various covariances among relatives are developed for the complete model which includes the configuration effects. Finally, the importance of this extension is discussed, primarily from the point of view of artificial selection.

II. Extension of the Gene Model

A random-mating population in equilibrium may be generated by multiplying the genotypic arrays for the various loci. Thus, for any number of alleles at each of two linked loci, the population may be represented as

$$\Pi = (\sum_{ij} p_i^1 p_j^1 A_i^1 A_j^1)(\sum_{kl} p_k^2 p_l^2 A_k^2 A_l^2)$$

$$= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 A_i^1 A_j^1 A_k^2 A_l^2,$$

where

$$\sum_{ij} p_i^1 p_j^1 A_i^1 A_j^1 = \text{genotypic array at locus 1,}$$

and

$$\sum_{kl} p_k^2 p_l^2 A_k^2 A_l^2 = \text{genotypic array at locus 2.}$$

Kempthorne (1957 for general reference) has utilized this representation for the elaboration of his gene model. Thus, if d_{ijkl} = genotypic value for $A_i^1 A_j^1 A_k^2 A_l^2$, such that

$$\sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 d_{ijkl} = 0,$$

then the Kempthorne model may be set out as follows:

$$d_{ijkl} = \alpha_i^1 + \alpha_j^1 + \alpha_k^2 + \alpha_l^2 + \delta_{ij}^1 + \delta_{kl}^2 + (\alpha\alpha)_{ik} + (\alpha\alpha)_{il} + (\alpha\alpha)_{jk} + (\alpha\alpha)_{jl}$$

$$+ (\alpha\delta)_{ikl} + (\alpha\delta)_{jkl} + (\delta\alpha)_{ijk} + (\delta\alpha)_{ijl} + (\delta\delta)_{ijkl},$$

where

α_u^a = additive genetic effect of the A_u^a allele,
δ_{uv}^a = dominance effect for the $A_u^a A_v^a$ genotype,
$(\alpha\alpha)_{ik}$ = additive × additive epistatic effect associated with genes A_i^1 and A_k^2,
$(\alpha\delta)_{ikl}$ = additive × dominance epistatic effect associated with the gene A_i^1 and the genotype $A_k^2 A_l^2$, and
$(\delta\delta)_{ijkl}$ = dominance × dominance epistatic effect associated with the genotypes $A_i^1 A_j^1$ and $A_k^2 A_l^2$.

This approach, however, does not permit the distinction between genotypic values for genotypes having the same set of genes but different chromosome arrangements. Hence, it is necessary to generate the population in such a way that the genotypes are represented by their chromosome constitution. This may be accomplished by simply squaring the chromosome array. For example, the population described above may be set out as follows:

$$\Pi = [\sum_{ik} p_i^1 p_k^2 (A_i^1 A_k^2)]^2$$

$$= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 (A_i^1 A_k^2)(A_j^1 A_l^2),$$

where

$$\sum_{ik} p_i^1 p_k^2 (A_i^1 A_k^2) = \text{chromosome array.}$$

The genotypic value of $(A_i^1 A_k^2)(A_j^1 A_l^2)$ may be designated as $d_{(ik)(jl)}$ such that

$$\sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 d_{(ik)(jl)} = 0.$$

In the remaining part of this section the argument will be concerned primarily with a random-mating population involving any number of alleles at each of two linked loci. Extensions to more than two loci will be briefly discussed at the end of the section.

(*a*) *Development of Model for Two Linked Loci*

The genotypic values in the two representations given above are related as follows:

$$d_{ijkl} = \tfrac{1}{2}(d_{(ik)(jl)} + d_{(il)(jk)}).$$

The inverse of this relationship leads to

$$\begin{aligned} d_{(ik)(jl)} &= d_{ijkl} + \tfrac{1}{2}(d_{(ik)(jl)} - d_{(il)(jk)}) \\ &= d_{ijkl} + c_{(ik)(jl)}, \end{aligned}$$

where

$$\begin{aligned} c_{(ik)(jl)} &= \tfrac{1}{2}(d_{(ik)(jl)} - d_{(il)(jk)}) \\ &= \text{effect due to the difference generated by the different chromosome configurations.} \end{aligned}$$

It is clear that

$$\begin{aligned} c_{(il)(jk)} &= \tfrac{1}{2}(d_{(il)(jk)} - d_{(ik)(jl)}) \\ &= -c_{(ik)(jl)}. \end{aligned}$$

Various summations involving the c values are of interest in subsequent arguments. These are:

(1)
$$\begin{aligned} c_{(ik)(..)} &= \Sigma p_j^1 p_l^2 c_{(ik)(jl)} \\ &= -\Sigma p_j^1 p_l^2 c_{(il)(jk)} \\ &= -c_{(i.)(.k)}. \end{aligned}$$

(2)
$$\begin{aligned} c_{(i.)(..)} &= \Sigma p_j^1 p_k^2 p_l^2 c_{(ik)(jl)} \\ &= -\Sigma p_j^1 p_k^2 p_l^2 c_{(il)(jk)} \\ &= -c_{(i.)(..)}. \end{aligned}$$

Hence

$$c_{(i.)(..)} = 0.$$

(3)
$$\begin{aligned} c_{(i.)(j.)} &= \Sigma p_k^2 p_l^2 \, c_{(ik)(jl)} \\ &= -\Sigma p_k^2 p_l^2 \, c_{(il)(jk)} \\ &= -c_{(i.)(j.)}. \end{aligned}$$

Hence

$$c_{(i.)(j.)} = 0.$$

The properties of $c_{(ik)(jl)}$, for the random-mating population, are as follows (where E denotes the expectation over i, j, k, and l):

$$\begin{aligned} E(c_{(ik)(jl)}) &= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 \, c_{(ik)(jl)} \\ &= \tfrac{1}{2}\left[\sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 \, (d_{(ik)(jl)} - d_{(il)(jk)})\right] \\ &= 0. \end{aligned}$$

Also

$$\begin{aligned} E(c_{(ik)(jl)})^2 &= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 \, (c_{(ik)(jl)})^2 \\ &= \sigma_C^2. \end{aligned}$$

Since $d_{(ik)(jl)} = d_{ijkl} + c_{(ik)(jl)}$, it is desirable to show that $c_{(ik)(jl)}$ is independent of d_{ijkl}. This may be accomplished in the following manner:

$$\begin{aligned} E(d_{ijkl} \cdot c_{(ik)(jl)}) &= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 \, [d_{ijkl} \cdot c_{(ik)(jl)}] \\ &= \tfrac{1}{4}\left[\sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 \, (d_{(ik)(jl)} + d_{(il)(jk)}) \, (d_{(ik)(jl)} - d_{(il)(jk)})\right] \\ &= \tfrac{1}{4}\left[\sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 \, (d_{(ik)(jl)}^2 - d_{(il)(jk)}^2)\right] \\ &= 0. \end{aligned}$$

Although $c_{(ik)(jl)}$ is independent of d_{ijkl}, and therefore independent of entire classes of elements in the gene model, it need not be independent of some of the individual elements. The following considers the expectations of cross-products of $c_{(ik)(jl)}$ with entire classes of effects as well as with individual component elements of the gene model.

(1) *Additive Effects*

$$\begin{aligned} E[c_{(ik)(jl)} \cdot (a_i^1 + a_j^1 + a_k^2 + a_l^2)] &= \sum_i p_i^1 \, c_{(i.)(..)} \, a_i^1 + \sum_j p_j^1 \, c_{(..)(j.)} a_j^1 \\ &\quad + \sum_k p_k^2 \, c_{(.k)(..)} a_k^2 + \sum_l p_l^2 \, c_{(..)(.l)} a_l^2 \\ &= 0, \end{aligned}$$

since

$$c_{(i.)(..)} = c_{(..)(j.)} = c_{(.k)(..)} = c_{(..)(.l)} = 0.$$

(2) *Dominance Effects*

$$\begin{aligned} E[c_{(ik)(jl)} \cdot (\delta_{ij}^1 + \delta_{kl}^2)] &= \sum_{ij} p_i^1 p_j^1 \, c_{(i.)(j.)} \, \delta_{ij}^1 + \sum_{kl} p_k^2 p_l^2 \, c_{(.k)(.l)} \, \delta_{kl}^2 \\ &= 0, \end{aligned}$$

since

$$c_{(i.)(j.)} = c_{(.k)(.l)} = 0.$$

(3) *Additive × Additive Effects*

$$\begin{aligned}E\{c_{(ik)(jl)}\cdot[(\alpha\alpha)_{ik}+(\alpha\alpha)_{il}+(\alpha\alpha)_{jk}+(\alpha\alpha)_{jl}]\}&\\ &= \sum_{ik} p_i^1 p_k^2\, c_{(ik)(..)}\,(\alpha\alpha)_{ik}+\sum_{il} p_i^1 p_l^2\, c_{(i.)(.l)}\,(\alpha\alpha)_{il}\\ &\quad+\sum_{jk} p_j^1 p_k^2\, c_{(.k)(j.)}\,(\alpha\alpha)_{jk}+\sum_{jl} p_j^1 p_l^2\, c_{(..)(jl)}\,(\alpha\alpha)_{jl}\\ &= [\sum_{ik} p_i^1 p_k^2\, c_{(ik)(..)}\,(\alpha\alpha)_{ik}-\sum_{il} p_i^1 p_l^2\, c_{(il)(..)}\,(\alpha\alpha)_{il}]\\ &\quad+[\sum_{jl} p_j^1 p_l^2\, c_{(..)(jl)}\,(\alpha\alpha)_{jl}-\sum_{jk} p_j^1 p_k^2\, c_{(..)(jk)}\,(\alpha\alpha)_{jk}]\\ &= 0.\end{aligned}$$

The fact that the difference within each bracket obviously equals zero does not imply that the individual terms within the brackets equal zero. Hence quantities of the type

$$\sum_{ik} p_i^1 p_k^2\, c_{(ik)(..)}\,(\alpha\alpha)_{ik},$$

need not equal zero.

(4) *Additive × Dominance Effects*

$$\begin{aligned}E\{c_{(ik)(jl)}\cdot[(\alpha\delta)_{ikl}+(\alpha\delta)_{jkl}]\} &= \sum_{ikl} p_i^1 p_k^2 p_l^2\, c_{(ik)(.l)}\,(\alpha\delta)_{ikl}+\sum_{jkl} p_j^1 p_k^2 p_l^2\, c_{(.k)(jl)}\,(\alpha\delta)_{jkl}\\ &= A+B\\ &= 0.\end{aligned}$$

The term A is equal to zero, since for each combination of alleles A_i^1, A_k^2, and A_l^2 there are two c configurations (due to the interchange of alleles at the A^2 locus) with the same frequency. These configurations are equal in magnitude but differ in sign, i.e.

$$c_{(ik)(.l)} = -c_{(il)(.k)}.$$

However, the interchange of A_k^2 and A_l^2 does not alter the value of $(\alpha\delta)_{ikl}$. Hence, the cross-product contributions involving A_i^1, A_k^2, and A_l^2 are

$$\begin{aligned}p_i^1 p_k^2 p_l^2[c_{(ik)(.l)}\,(\alpha\delta)_{ikl}+c_{(il)(.k)}\,(\alpha\delta)_{ilk}] &= p_i^1 p_k^2 p_l^2[c_{(ik)(.l)}\,(\alpha\delta)_{ikl}-c_{(ik)(.l)}\,(\alpha\delta)_{ikl}]\\ &= 0.\end{aligned}$$

Similarly, the term B is equal to zero.

(5) *Dominance × Additive Effects*

An argument similar to that given in (4) is applicable.

(6) *Dominance × Dominance Effects*

$$\begin{aligned}E[c_{(ik)(jl)}\cdot(\delta\delta)_{ijkl}] &= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2\, c_{(ik)(jl)}\cdot(\delta\delta)_{ijkl}\\ &= 0.\end{aligned}$$

That this summation equals zero can be seen by considering all possible permutations of each combination of A_i^1, A_j^1, A_k^2, and A_l^2, which are generated by inter-

changing A_i^1 and A_j^1 as well as A_k^2 and A_l^2. Such operations may change the value of the c values but do not alter the values of the $(\delta\delta)$ terms. Hence, the cross-product contribution of the combination $A_i^1A_j^1A_k^2A_l^2$ is

$$p_i^1p_j^1p_k^2p_l^2[c_{(ik)(jl)}(\delta\delta)_{ijkl}+c_{(jk)(il)}\ (\delta\delta)_{jikl}+c_{(il)(jk)}\ (\delta\delta)_{iljk}+c_{(jl)(ik)}\ (\delta\delta)_{jilk}]$$
$$= p_i^1p_j^1p_k^2p_l^2[2c_{(ik)(jl)}-2c_{(ik)(jl)}](\delta\delta)_{ijkl}$$
$$= 0.$$

In summary then, the gene model for the genotype $(A_i^1A_k^2)\ (A_j^1A_l^2)$ is

$$d_{(ik)(jl)} = a_i^1+a_j^1+a_k^2+a_l^2+\delta_{ij}^1+\delta_{kl}^2+(a\alpha)_{ik}+(a\alpha)_{il}+(a\alpha)_{jk}+(a\alpha)_{jl}$$
$$+(a\delta)_{ikl}+(a\delta)_{jkl}+(\delta a)_{ijk}+(\delta a)_{ijl}+(\delta\delta)_{ijkl}+c_{(ik)(jl)},$$

where all effects are independent of each other except that the individual $(a\alpha)$'s are not independent of $c_{(ik)(jl)}$. The total genotypic variance may be partitioned as:

$$\sigma^2 = \sigma_A^2+\sigma_D^2+\sigma_{AA}^2+\sigma_{AD}^2+\sigma_{DD}^2+\sigma_C^2.$$

(*b*) *Extension of Model to more than Two Linked Loci*

In extending the theory to more than two linked loci, the first problem is to determine the number of different genotypes which are possible by permuting the two alleles at each of an arbitrary number of loci.

For the ath locus with alleles A_i^a and A_j^a the arrangements may be generated by a permutation group of order two, i.e. $[G = I, (ij)]$. Hence, for n loci, all possible arrangements may be obtained by application of an Abelian permutation group of order 2^n which results as a direct product of the n permutation groups of order two.

However, not all gene–chromosome arrangements give rise to different genotypes, since genotypes are invariant to permutation of chromosomes. In fact, the 2^n permutations may be paired such that in one member of a pair, alleles at all loci are interchanged which are not permuted in the other. Such pairs of permutations generate equivalent arrangements, one of which may be derived from the other by chromosome interchange. Hence, the number of different genotypes is 2^{n-1}.

The following tabulation presents the simplest illustration of the above argument. It gives the permutation group, gene–chromosome arrangements, and genotypes which are possible for two linked heterozygous loci.

Permutation Group	Gene–Chromosome Arrangements	Genotypes
I	$(A_i^1A_k^2)\ (A_j^1A_l^2)$	$(A_i^1A_k^2)\ (A_j^1A_l^2) = (A_j^1A_l^2)\ (A_i^1A_k^2)$
(ij)	$(A_j^1A_k^2)\ (A_i^1A_l^2)$	$(A_j^1A_k^2)\ (A_i^1A_l^2) = (A_i^1A_l^2)\ (A_j^1A_k^2)$
(kl)	$(A_i^1A_l^2)\ (A_j^1A_k^2)$	
$(ij)\ (kl)$	$(A_j^1A_l^2)\ (A_i^1A_k^2)$	

Finally, there are a total of 2^{n-1} configuration constants for genotypes heterozygous for the same set of n linked loci, since a constant is associated with each

genotype. However, since these constants must sum to zero, the number of independent effects is $2^{n-1}-1$.

As an illustrative example, consider the situation which arises for three linked loci. The four possible genotypes are: $(A_i^1A_k^2A_m^3)$ $(A_j^1A_l^2A_n^3)$, $(A_j^1A_k^2A_m^3)$ $(A_i^1A_l^2A_n^3)$, $(A_i^1A_l^2A_m^3)$ $(A_j^1A_k^2A_n^3)$, and $(A_i^1A_k^2A_n^3)$ $(A_j^1A_l^2A_m^3)$.

Since

$$d_{ijklmn} = \tfrac{1}{4}[d_{(ikm)(jln)}+d_{(jkm)(iln)}+d_{(ilm)(jkn)}+d_{(ikn)(jlm)}],$$

then

$$d_{(ikm)(jln)} = d_{ijklmn}+c_{(ikm)(jln)},$$
$$d_{(jkm)(iln)} = d_{ijklmn}+c_{(ikm)(jln)}, \text{ etc.},$$

where

$$c_{(ikm)(jln)} = \tfrac{1}{4}[(d_{(ikm)(jln)}-d_{(jkm)(iln)})+(d_{(ikm)(jln)}-d_{(ilm)(jkn)}) +(d_{(ikm)(jln)}-d_{(ikn)(jlm)})],$$
$$c_{(jkm)(iln)} = \tfrac{1}{4}[(d_{(jkm)(iln)}-d_{(ikm)(jln)})+(d_{(jkm)(iln)}-d_{(ilm)(jkn)}) +(d_{(jkm)(iln)}-d_{(ikn)(jlm)})],$$
$$c_{(ilm)(jkn)} = \tfrac{1}{4}[(d_{(ilm)(jkn)}-d_{(ikm)(jln)})+(d_{(ilm)(jkn)}-d_{(jkm)(iln)}) +(d_{(ilm)(jkn)}-d_{(ikn)(jlm)})],$$

and

$$c_{(ikn)(jlm)} = \tfrac{1}{4}[(d_{(ikn)(jlm)}-d_{(ikm)(jln)})+(d_{(ikn)(jlm)}-d_{(jkm)(iln)}) +(d_{(ikn)(jlm)}-d_{(ilm)(jkn)})].$$

However, since the following linear restriction holds, there are only three independent constants:

$$c_{(ikm)(jln)}+c_{(jkm)(iln)}+c_{(ilm)(jkn)}+c_{(ikn)(jlm)} = 0.$$

From this brief discussion, it is clear that it is conceptually possible to define and enumerate configuration effects for any number of linked loci.

III. Expectations of Covariances among Relatives

Estimation of the additive variance component is essential if the permanent gains from selection are to be predicted. In the past, various covariances among relatives have been used to make this estimation. The covariances of interest are: parent–offspring covariance, designated as Cov(PO); half-sib covariance, designated as Cov(HS); and full-sib covariance, designated as Cov(FS).

The objective of this section is to develop the expectations of these covariances for a two-locus gene model which is generalized to include the following details:

(i) any number of alleles at each locus;

(ii) any system of dominance and epistatic parameters;

(iii) recombination values which may be different for the two sexes, i.e.

y_m = recombination value for males, and
y_f = recombination value for females; and

(iv) gene–chromosome configuration effects.

Since the expectations of the covariances have been derived for a gene model which is generalized for all but the inclusion of configuration effects (Griffing 1960*b*), the purpose, here, is to see how these configuration effects disturb the covariances, and hence the estimation of variance components from these covariances.

The parent–offspring covariance may be defined as the expected cross-product of the genotypic value of an arbitrary parent individual and the genotypic mean of the half-sib array associated with the parent individual. If configuration constants are not considered, it can be shown that linkage does not disturb Cov(PO). However, when these effects are included, not only does linkage disturb this covariance but the covariance for males may be different from that for females. Therefore, they must be treated separately. Consider first the male covariance which may be designated as $\mathrm{Cov}_{(m)}(\mathrm{PO})$.

An arbitrary male $(A_i^1A_k^2)$ $(A_j^1A_l^2)$ produces the following gametic array:

$$\{[(1-y_m)/2]\,(A_i^1A_k^2+A_j^1A_l^2)+(y_m/2)\,(A_i^1A_l^2+A_j^1A_k^2)\}.$$

The total female gametic array for the random-mating population is

$$\sum_{rt} p_r^1 p_t^2\,(A_r^1A_t^2).$$

Therefore, the male half-sib family mean is

$$\begin{aligned} h_{(ik,\ jl)\,(\ldots,\ \ldots)} &= [(1-y_m)/2]\sum_{rt} p_r^1p_t^2\,[d_{(ik)(rt)}+d_{(jl)(rt)}] \\ &\quad +(y_m/2)\sum_{rt} p_r^1p_t^2\,[d_{(il)(rt)}+d_{(jk)(rt)}] \\ &= [(1-y_m)/2]\,(d_{(ik)(\ldots)}+d_{(jl)(\ldots)})+(y_m/2)\,(d_{(il)(\ldots)}+d_{(jk)(\ldots)}). \end{aligned}$$

The male parent–offspring covariance is then

$$\begin{aligned} \mathrm{Cov}_{(m)}(\mathrm{PO}) &= \sum_{ijkl} p_i^1p_j^1p_k^2p_l^2(h_{(ik,\ jl)(\ldots,\ \ldots)})\,(d_{(ik)(jl)}) \\ &= \tfrac{1}{2}\sigma_A^2+\tfrac{1}{4}\sigma_{AA}^2+(1-2y_m)\sum_{ik} p_i^1p_k^2(c_{(ik)(\ldots)})^2 \\ &\quad +2(1-y_m)\sum_{ik} p_i^1p_k^2(c_{(ik)(\ldots)})\,(\alpha\alpha)_{ik}. \end{aligned}$$

Similarly, the female parent–offspring covariance is

$$\begin{aligned} \mathrm{Cov}_{(f)}(\mathrm{PO}) &= \tfrac{1}{2}\sigma_A^2+\tfrac{1}{4}\sigma_{AA}^2+(1-2y_f)\sum_{ik} p_i^1p_k^2[c_{(ik)(\ldots)}]^2 \\ &\quad +2(1-y_f)\sum_{ik} p_i^1p_k^2[c_{(ik)(\ldots)}]\,(\alpha\alpha)_{ik}. \end{aligned}$$

The half-sib covariance may be defined as the expectation of the squares of the half-sib family means. Again, male and female half-sib covariances may be different. Consider, first, the derivation of the male half-sib covariance.

Since the half-sib family mean for an arbitrary male $(A_i^1 A_k^2)$ $(A_j^1 A_l^2)$ has been given, it is obvious that the male covariance of half-sibs is

$$\begin{aligned}\mathrm{Cov}_{(m)}(\mathrm{HS}) &= \sum_{ijkl} p_i^1 p_j^1 p_k^2 p_l^2 (h_{(ik,\ jl)(\ldots,\ \ldots)})^2 \\ &= \tfrac{1}{4}\sigma_A^2 + [(1+\delta_m)/16]\sigma_{AA}^2 + [(1+\delta_m)/4] \sum_{ik} p_i^1 p_k^2 (c_{(ik)(\ldots)})^2 \\ &\quad + [(1+\delta_m)/2] \sum_{ik} p_i^1 p_k^2 (c_{(ik)(\ldots)})\ (\alpha\alpha)_{ik},\end{aligned}$$

where

$$\delta_m = (1-2y_m)^2.$$

Likewise, the female half-sib covariance is

$$\begin{aligned}\mathrm{Cov}_{(f)}(\mathrm{HS}) &= \sum_{rstu} p_r^1 p_s^1 p_t^2 p_u^2 (h_{(\ldots,\ \ldots)(rt,\ su)})^2 \\ &= \tfrac{1}{4}\sigma_A^2 + [(1+\delta_f)/16]\sigma_{AA}^2 + [(1+\delta_f)/4] \sum_{rt} p_r^1 p_t^2 (c_{(\ldots)(rt)})^2 \\ &\quad + [(1+\delta_f)/2] \sum_{rt} p_r^1 p_t^2 (c_{(\ldots)(rt)})\ (\alpha\alpha)_{rt},\end{aligned}$$

where

$$\delta_f = (1-2y_f)^2.$$

Finally, the full-sib covariance may be defined as the expected value of the squares of the full-sib means. Consider, now, the evaluation of Cov(FS).

The mean of the full-sib array which results from the cross between an arbitrary male $(A_i^1 A_k^2)$ $(A_j^1 A_l^2)$ and an arbitrary female $(A_r^1 A_t^2)$ $(A_s^1 A_u^2)$ is

$$\begin{aligned}h_{(ik,\ jl)(rt,\ su)} = \{&[(1-y_m)/2][(1-y_f)/2](d_{(ik)(rt)} + d_{(ik)(su)} + d_{(jl)(rt)} + d_{(jl)(su)}) \\ &+ [(1-y_m)/2](y_f/2)(d_{(ik)(ru)} + d_{(ik)(st)} + d_{(jl)(ru)} + d_{(jl)(st)}) \\ &+ (y_m/2)[(1-y_f)/2](d_{(il)(rt)} + d_{(il)(su)} + d_{(jk)(rt)} + d_{(jk)(su)}) \\ &+ (y_m/2)(y_f/2)(d_{(il)(ru)} + d_{(il)(st)} + d_{(jk)(ru)} + d_{(jk)(st)})\}.\end{aligned}$$

The covariance of full-sibs is then

$$\begin{aligned}\mathrm{Cov}(\mathrm{FS}) &= \sum_{ijklrstu} p_i^1 p_j^1 p_r^1 p_s^1 p_k^2 p_l^2 p_t^2 p_u^2 (h_{(ik,\ jl)(rt,\ su)})^2 \\ &= \tfrac{1}{2}\sigma_A^2 + \tfrac{1}{4}\sigma_D^2 + \{\tfrac{1}{4} + [(\delta_f+\delta_m)/16]\}\sigma_{AA}^2 + \{\tfrac{1}{8} + [(\delta_f+\delta_m)/16]\}\sigma_{AD}^2 \\ &\quad + [\tfrac{1}{16}(1+\delta_f)(1+\delta_m)]\sigma_{DD}^2 + [\tfrac{1}{16}(1+\delta_f)(1+\delta_m)]\sigma_C^2 \\ &\quad + [\tfrac{1}{8}(1+\delta_f)(1+\delta_m) + 2(y_m)(y_f)(1-y_m)(1-y_f)] \sum_{ik} p_i^1 p_k^2 (c_{(ik)(\ldots)})^2 \\ &\quad + \{[y_f(1-y_f)/4](1+\delta_m) + [y_m(1-y_m)/4](1+\delta_f)\} \\ &\quad \times \{ \sum_{irk} p_i^1 p_r^1 p_k^2 (c_{(ik)(r.)})^2 + \sum_{ikt} p_i^1 p_k^2 p_t^2 (c_{(ik)(.t)})^2\} \\ &\quad + [\tfrac{1}{4}(1+\delta_f)(1+\delta_m) - 4(y_m)(y_f)(1-y_m)(1-y_f)] \sum_{ik} p_i^1 p_k^2 [c_{(ik)(\ldots)}](\alpha\alpha)_{ik}.\end{aligned}$$

Assuming (i) epistatic interactions involving three or more loci are negligible, and (ii) configuration effects are absent, it has been shown (Griffing 1960*b*) that σ_A^2 and σ_{AA}^2 can be estimated as follows:

$$\hat{\sigma}_A^2 = \frac{\mathrm{Cov}(\mathrm{PO})[2\bar{y}_m(1-\bar{y}_m) - 1] + 2[\mathrm{Cov}_{(m)}(\mathrm{HS})]}{\bar{y}_m(1-\bar{y}_m)},$$

and

$$\hat{\sigma}^2_{AA} = \frac{2\{\mathrm{Cov(PO)} - 2[\mathrm{Cov}_{(m)}(\mathrm{HS})]\}}{\bar{y}_m(1-\bar{y}_m)},$$

where $\bar{y}_m$ represents the recombination value averaged over all possible pairs of active loci as measured in the male sex.

However, it is now clear that if the configuration effects are taken into consideration, these variance component estimates are no longer unbiased.

IV. Discussion

The question of how far and in what way the inclusion of configuration effects may disturb prediction theory is discussed below.

Earlier it was pointed out that there are at least two ways in which configuration effects may be generated. First, the physical configuration of the genes may induce the position-effect phenomenon. Second, linkage may give rise to configuration effects if the various gene–chromosome arrangements are evaluated on the basis of their inbred progeny. These different sources of disturbance are discussed separately.

A discussion of the position-effect phenomenon necessitates a brief consideration of the modern concept of the "gene". This concept postulates that the chromosome may be divided into functional regions each of which controls a specific biological activity. These functional regions may each contain numerous mutational and recombinational sites. In some cases it has been shown that linear linkage maps may be obtained from intralocus recombination data. Mutations are assigned to functional regions on the basis of the position-effect criterion, the site of the mutation being the muton (see Benzer 1957). Thus, mutations exhibiting position effect are assigned to the same functional region and those that do not exhibit this effect are assigned to different regions.

From this concept of the gene two points need to be considered with regard to the importance of the position-effect phenomenon in prediction theory. First, there are at least two possibilities in the choice of a basic hereditary unit on which the selection theory rests. Second, position effects are usually generated by mutations which are very close together, i.e. mutations in the same functional region.

The choice of a basic hereditary unit has been discussed previously (Griffing 1960*a*):

"There are at least two methods of representing the genetic situation at a complex locus. To illustrate, consider a locus which has a simplified structure consisting of only two genetic conditions (mutant and normal) at each of two mutational sites. In the first method, the locus can be subdivided into two subloci, one for each of the mutational sites. This approach yields two sets of alleles, each set being the genetic alternatives at each sublocus. In this case, the gene model for quantitative inheritance must be extended to accommodate position effect which may occur between alleles at different subloci. This, so far, has never been done.

The alternative method is to consider the overall locus as the basic entity, and to regard all possible genetic structures at this locus as the set of multiple alleles. Thus, in the simplified example, the four possible gene states are $(+\ +)$, $(m_1\ +)$, $(+\ m_2)$, and $(m_1\ m_2)$. These, then, would be regarded as the alleles of the locus. Such a representation avoids the introduction of intralocus position effect because complexities such as the *cis–trans* relations would be absorbed

in the dominance parameters. However, a resultant complication of this approach is that mutation of alleles as defined above includes both point mutation in its conventional sense and intralocus recombination. For example, recombination between mutational sites arranged in a *trans* configuration, $m_1 +/+ m_2$, results in non-parental locus types, $(+ +)$ and $(m_1 m_2)$. It is of course clear that the frequency of such intralocus recombination is low compared with the frequency of recombination between genes at different loci. Hence, it would appear that with the alleles defined as above, the contribution of locus mutation (point mutation and intralocus recombination) would be negligible in most theoretical plant and animal breeding studies."

If, then, it is satisfactory to regard the entire functional region as the basic unit of inheritance, the problem of position effects disappears. Such a solution seems appropriate for short-term selection theory. However, it might not be completely satisfactory for a theory pertaining to selection sustained for a very long time. In this case, it may be best to consider the mutational site as the basic hereditary unit, and hence the disturbance due to position effects should be examined.

Assuming, then, that the basic hereditary unit is the muton, the following argument considers the relative frequency, and hence the importance, of the position-effect phenomenon as it occurs among all possible pairwise combinations of mutons. It is assumed that the genotypic variability associated with the given quantitative variable is controlled by mutations, each of small effect at many mutons which are scattered at random over the chromosome complement.

The argument is: (i) the position-effect phenomenon is generated only by mutons in the same functional region and not by mutons in different regions, and (ii) in general, as the number of active regions increases, the frequency of pairwise combinations of mutational sites in different functional regions increases relative to the frequency of pairwise combinations of sites in the same functional region. Hence, when considering all possible pairs of mutons, the phenomenon becomes increasingly rare as the number of active functional regions increases.

This argument can be set out more rigorously as follows: Let there be n mutational sites (mutons) in each of N functional regions. Thus, there are a total of Nn sites. Of the total number of pairwise combinations $\binom{Nn}{2}$, there are $N\binom{n}{2}$ combinations of mutons in which both members of the pair are in the same functional region and hence may give rise to position effects. The remaining combinations, which number $\binom{N}{2}n^2$, have one muton in one functional region and the other in a different region. Hence they cannot give rise to positional effects. The relative proportion of pairs of sites in which position effects cannot occur is

$$\frac{\binom{N}{2}n^2}{\binom{Nn}{2}} = \frac{n[1-(1/N)]}{[n-(1/N)]}.$$

Therefore, as the number of functional regions becomes large this proportion approximates one, irrespective of the number of mutons per functional unit.

The conclusion is that, even if position effects are widespread, the disturbance they cause to selection theory is negligible if the assumption holds that the genetic

variability is generated by genetic alternatives of small effect at many mutons (i.e. by genes of small effect at many loci).

Consider now the second phenomenon in which the individual is evaluated by its inbred progeny. In this case, the configuration effects are due to linkage and are not confined only to genes in close proximity but to all linked genes which do not show independent segregation.

Since these effects are directly a function of recombination values, the argument pertaining to the estimation of the average recombination value, as given by Griffing (1960*b*), is appropriate. It was shown that a fairly accurate estimate of the recombination value averaged over all possible pairs of active loci may be obtained from the formula

$$\bar{y} = (r-1)/2r,$$

where r = recombination index (Darlington 1958), i.e. the sum of the haploid number of chromosomes and the average number of chiasmata per cell.

This formula implies, roughly, that, of all pairwise combinations of loci, the relative proportion which segregates independently is $(r-1)/r$. Hence, as the recombination index increases, this value rapidly approximates one.

Again, this argument is subject to various assumptions and approximations, but nevertheless, it appears that if the haploid chromosome number is five or more (i.e. $r > 10$, because at least one chiasma per bivalent is obligatory), the disturbance due to linkage is probably not great. However, there are certainly instances in which configuration effects cannot be completely ignored. These include (1) cases in which crossing over is greatly reduced or non-existent (as in male *Drosophila*) in an organism which has a low chromosome number, and (2) cases in which interest centres on the manipulation of the gene content in a small number of chromosome pairs.

Finally, it is necessary to point out just where the configuration effects cause a disturbance to the prediction theory, if, in fact, they are appreciable.

It was stated previously that to predict *permanent* gains from selection, it is necessary to estimate the additive genetic variance. Configuration effects may then lead to biased estimates of this variance component. However, the estimation of the additive genetic variance component is not necessary to predict the *immediate* gains from selection. It can be shown that for both configuration-effect phenomena this prediction may be made directly from certain covariances. Since there is no theoretical difficulty in estimating these covariances, irrespective of the presence or absence of configuration effects, there is no bias in the estimation of immediate gains from artificial selection (ignoring, of course, the effects of natural selection).

It is clear, however, that the immediate gains may not be entirely sustained on relaxation of selection, and it is the gains which are retained after relaxation that are termed, here, the permanent gains. It is the prediction of these gains that may be biased by the presence of the configuration effects.

V. Acknowledgment

I am indebted to my colleague, Dr. B. D. H. Latter, for his valuable suggestions.

VI. References

Benzer, S. (1957).—The elementary units of heredity. In "The Chemical Basis of Heredity". (Eds. W. D. McElroy and B. Glass.) (The Johns Hopkins Press: Baltimore.)

Darlington, C. D. (1958).—"Evolution of Genetic Systems." 2nd Ed. (Oliver and Boyd: London.)

Griffing, B. (1960*a*).—Theoretical consequences of truncation selection based on the individual phenotype. *Aust. J. Biol. Sci.* **13**: 307–43.

Griffing, B. (1960*b*).—Accommodation of linkage in mass selection theory. *Aust. J. Biol. Sci.* **13**: 501–26.

Kempthorne, O. (1957).—"An Introduction to Genetic Statistics." (John Wiley and Sons, Inc.: New York.)

Sturtevant, A. (1925).—The effects of unequal crossing over at the Bar locus in *Drosophila*. *Genetics* **10**: 117–47.