THE EFFECT OF INBREEDING ON THE VARIATION DUE TO RECESSIVE GENES *

ALAN ROBERTSON
Animal Breeding and Genetics Research Organization, Edinburgh, Scotland

Received June 25, 1951

IN his classical treatment of inbreeding, WRIGHT (1921) developed the concept of the inbreeding coefficient F, which he defined as the correlation between the genetic constitution of the gametes in the uniting egg and sperm. This is directly related to the heterozygosity remaining in the population which is equal to $1-F$ times the heterozygosity at the start of inbreeding. If a number of inbred lines are made without selection from a random breeding population, the genetic variance due to genes which act additively increases between lines as $2F$ and decreases within lines as $1-F$. (If inbreeding is rapid, the value $1-F$ for the genetic variance within lines is not adequate, the correct expression being $1+\overline{F}-2\overline{F}$, where $\overline{F}$ is the inbreeding coefficient of the hypothetical progeny produced by random mating within lines in the present generation and $2\overline{F}$ is correct for the variance between lines.) For genes which do not act additively, there is not the same correspondence between heterozygosity and variance and the above relationships do not hold. As we know little about the dominance relationships of the genes controlling continuous variation, it seemed desirable to investigate theoretically the effect of inbreeding on the variation due to genes which are completely recessive and to genes which show overdominance. Particular attention is given to the case in which the recessive (or quasi-recessive) is at low frequency as this is the most probable situation in natural populations.

We shall deal first with continued full-sib mating in which the results can be worked out by simple, if rather laborious, arithmetic and where the process can be most easily visualised. The more general situation of slow inbreeding in lines of constant breeding size requires more sophisticated mathematics and the details of the derivations are given in an appendix. The two methods are in good agreement.

CONTINUED FULL-SIB MATING

This system of mating can be treated most simply by the method of mating types, originally used by JENNINGS (1916) and subsequently with the help of matrix theory by HALDANE (1937) and by FISHER (1949). If only two alleles, A and a, are present in the population at a locus, there are six possible types of matings. The relative frequencies of matings of different types in any generation can be calculated from the frequencies in the previous generation

* Part of the cost of the accompanying mathematical formulae has been paid by the GALTON and MENDEL MEMORIAL FUND.

on the assumption that the offspring are mated at random. For instance, matings Aa × aa will give offspring ½ Aa, ½ aa and if these are mated at random, one quarter of the matings will be of the type Aa × Aa, one half Aa × aa and one quarter aa × aa. The equations giving the frequencies of the six types in terms of the frequencies in the preceding generation are shown schematically in table 1 in which y_0, for instance, signifies the frequency of AA × Aa matings in the zero generation.

Reading horizontally, we have, for instance, $u_1 = \frac{1}{4} y_0 + z_0 + \frac{1}{4} u_0 + \frac{1}{4} v_0$. If a is completely recessive to A and the phenotypic value of AA, Aa is taken as zero and of aa as unity, the genetic variance within the progeny of Aa × Aa and Aa × aa matings is 3/16 and 1/4 respectively, the other mating types having no variation within their progeny. (Throughout the paper, "genetic variance" will be used in the sense of all variance due to gene segregation.) The average variance within lines (each line being in this case a single mating) is 3/16 u + 1/4 v. The genetic variance between lines is easily calculated, as the expected value for the progeny is zero for all types except those represented by u, v, and w for which it is ¼, ½ and 1 respectively.

TABLE 1

Frequencies of the various mating types.

	AA × AA x_0	AA × Aa y_0	AA × aa z_0	Aa × Aa u_0	Aa × aa v_0	aa × aa w_0
x_1	1	¼		1/16		
y_1		½		¼		
z_1				⅛		
u_1		¼	1	¼	¼	
v_1				¼	½	
w_1				1/16	¼	1

In the computations, we take as x_0, y_0, z_0, etc. the values for a random-bred population and evaluate the set x_1, y_1, z_1, etc., and so on. Figure 1 shows the frequencies of mating types other than AA × AA for the case when $q_a = 0.1$. The immediate effect of the inbreeding is to cause y to decrease and u and v to increase. After about 6 generations, the frequencies of all types of matings except AA × AA and aa × aa become practically constant relative to one another and then decline to zero when inbreeding is complete. At that point, x = 0.90 and w = 0.10, as all lines become homozygous for either A or a. The variance within lines increases considerably in the first generation (F = 0.25) to 2.99 times its random breeding value, remains fairly stationary for two further generations, and then declines. The variance between lines increases continually as the inbreeding progresses, the rise being almost linear for the first six generations (fig. 2). Consideration of the first generation only shows that the variance within lines will only increase above its random breeding value if the frequency, q, of the recessive gene is below 0.47. At low gene frequencies, at the start almost all the a genes will be carried in matings AA × Aa which will have frequency approximately 4q, the frequencies of other matings in which heterozygotes take part being in higher powers of q. Thus

all the matings, except those of type AA × AA, will derive from an initial entry of y_0 as 4q. It follows that the shape of the curves showing the effect of inbreeding on variation will become independent of q as q decreases, as all entries will be multiples of q. Computation shows that in this case, the variance within matings reaches its maximum value of 0.207q after three generations of inbreeding (F = 0.50). Under random mating conditions, matings of type Aa × Aa will make the major contribution to the variance within mat-

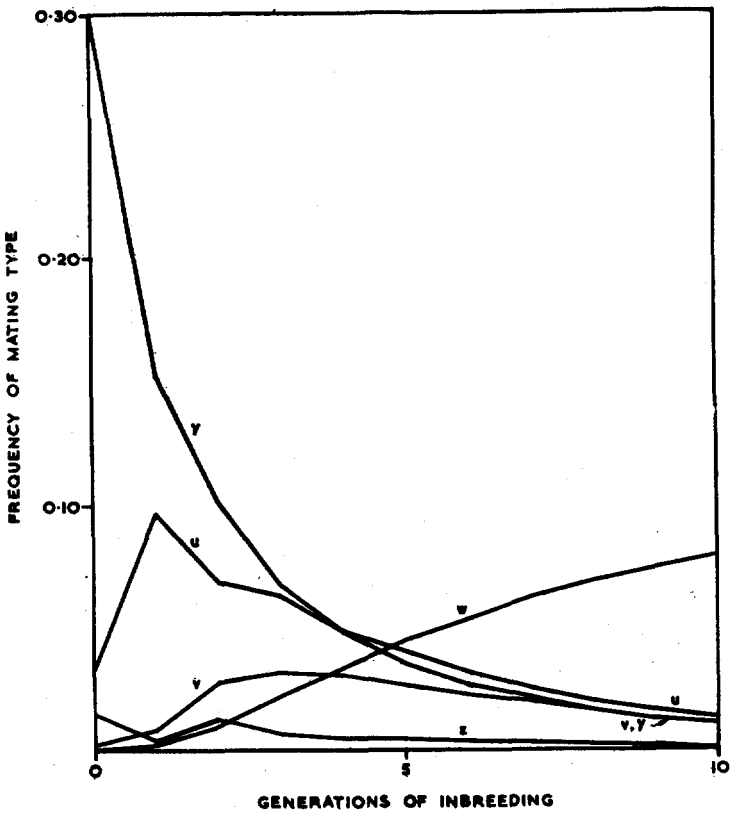

FIGURE 1.—Frequencies of various mating types (except AA × AA) with continued full-sib mating. The mating types are designated as: y (AA × Aa), z (AA × aa), u (Aa × Aa), v (Aa × aa), and w (aa × aa).

ings with frequency $4q^2(1-q)^2 \sim 4q^2$ and variance 3/16, giving an average variance of $\frac{3}{4}\ q^2$. At its maximum, the variance within matings is therefore $\frac{0.207q}{0.75q^2} = \frac{0.276}{q}$ times its random breeding value. Figure 3 shows the curves for the different variances when q is very low, which are not very different from those in figure 2 for q = 0.10.

The total genetic variance rises continuously as the inbreeding progresses. When complete homozygosis is reached, the total variance is $q(1-q)$, q lines having phenotypic value unity and $1-q$ having value zero. With random mating, when a fraction q^2 have genotype aa and phenotype value one and

the rest have value zero, the total variance is $q^2(1-q^2)$. Of this, the additively genetic component, the portion usually detected in genetic analyses such as heritability studies, is only $2q^3(1-q)$. When inbreeding is complete, the total variance is equal to $1/q(1-q)$ times its random breeding value and $1/2\,q^2$ times the part of the variation that can usually be detected in the absence of inbreeding. For genes which act strictly additively, the total variance at complete homozygosis is twice the total variance under random breeding conditions, the latter being, of course, equal to the additive component.

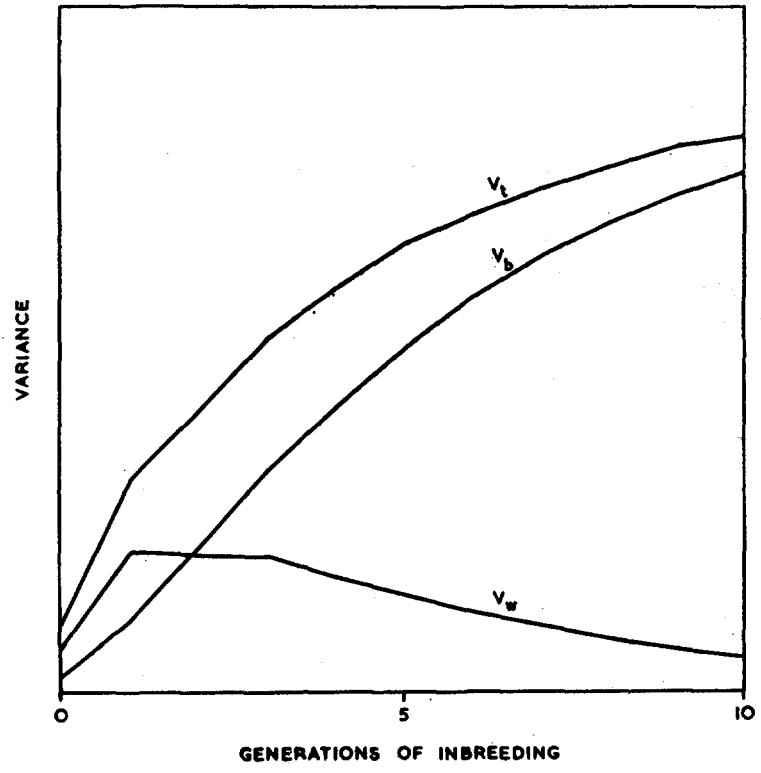

FIGURE 2.—Total variance (V_t), variance within lines (V_w), and variance between lines (V_b) with continued full-sib mating and $q_0 = 0.1$.

INBREEDING IN LINES OF CONSTANT BREEDING SIZE

In the more general case with slow inbreeding, in which each line is of constant breeding size, it is more convenient to consider the inbreeding from the point of view of the change of gene frequency in the several lines. Starting from a population with a given gene frequency, the frequencies in the lines, as inbreeding progresses, gradually scatter further and further from the original value due to sampling (the phenomenon which in discussions of evolution is often called the " Sewall Wright drift "). The frequency averaged over the whole population of lines remains the same as that in the original

random breeding population, if there has been no selection. When inbreeding is complete, the gene frequency in each line is either 0 or 1. The variance within and between lines can then be related to the distribution of the gene frequency in the several lines. Within a line in which the gene frequency is q_1, the genetic variance is $q_1^2(1-q_1^2)$ so that the average value of the genetic variance within lines is $\mu_2 - \mu_4$, where the μ's are the moments of the q distribution about zero. By a similar argument, the genetic variance between lines

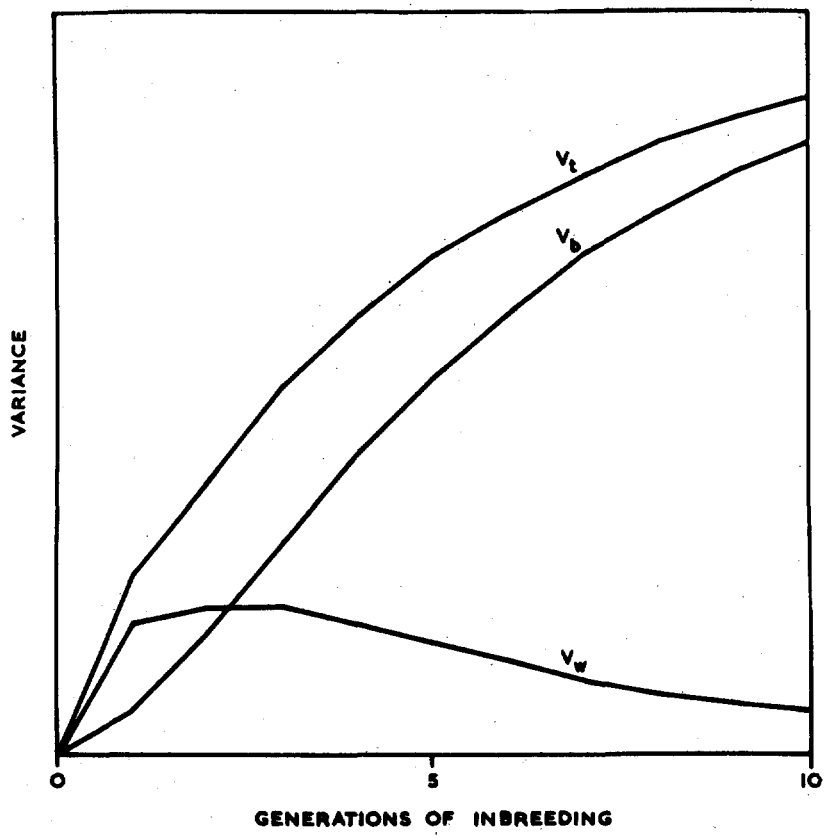

FIGURE 3.—Total variance (V_t), variance within lines (V_w), and variance between lines (V_b) with continued full-sib mating and q_0 very small.

is the variance of q_1^2, $\mu_4 - \mu_2^2$ and the total genetic variance is $\mu_2 - \mu_2^2$. The evaluation of general expressions for the moments, and therefore of the variances, as inbreeding progresses, depends on matrix theory and is given in an appendix. The within line variance is given by

$$V_w = a(1-F) + b(1-F)^3 + c(1-F)^6$$

where $a = 0.8\,q(1-q)$

$b = -q(1-q)(1-2q)$

$c = 0.2q(1-q) - q^2(1-q)^2$.

When $F = 0$, $V_w = a + b + c = q^2(1 - q^2)$ and when $F = 1$, $V_w = 0$. The change of V_w when F is small can be judged from the coefficient of F in the expansion of the above equation, which on simplification becomes $q(1 - q)(1 - 6q^2)$. V_w will then only pass through a maximum if $1 - 6q^2 > 0$, *i.e.*, $q < 0.41$ and will otherwise decline continuously. As F increases, the last two terms in V_w decrease in importance until V_w becomes proportional to $1 - F$. When q is small, V_w becomes $q\ [0.8(1 - F) - (1 - F)^3 + 0.2(1 - F)^6]$, and rises to a

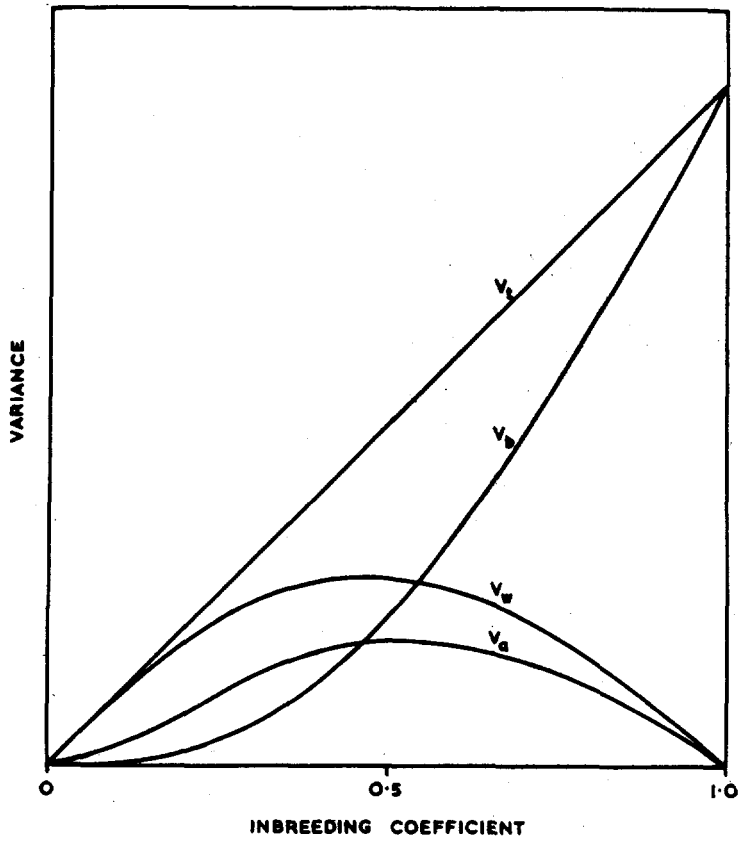

FIGURE 4.—Additive variance within lines (V_a), total variance (V_t), variance within lines (V_w) and variance between lines (V_b) with random mating in a population of constant size with q_a very small.

maximum of 0.280q, when $F = 0.46$, compared to $q^2(1 - q^2)$ in the random bred population. The maximum value is then roughly $\frac{0.280}{q}$ times the random-breeding value, in good agreement with the value of $\frac{0.276}{q}$ obtained from continual full-sib mating.

The additive component of the variance within lines, the component detectable by such techniques as parent-offspring regression, is for a given line

$2q_1^3(1-q_1)$ so that the average value of this, V_a, is $2(\mu_3-\mu_4)$. In the above terminology this is given by $.75a(1-F)+b(1-F)^3+2c(1-F)^6$.

It is not possible to give simple formulae for the between line or total variance except when q is small. Then, expanding V_b as a power of F, the term with the lowest power of y is $3F^3q$. The total variance is then Fq. Figures 4 and 5 show the behaviour of V_t, V_b, and V_w when q is very small and when q = 0.10. The curves show clearly the main features of the effect of inbreeding on the variance. V_w and V_a increase to a maximum when F = 0.4

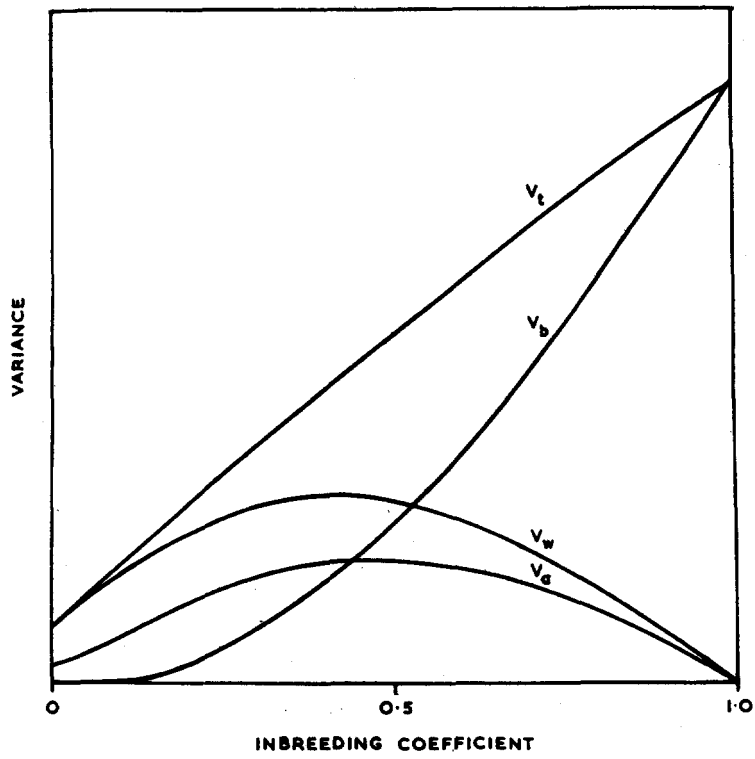

FIGURE 5.—Additive variance within lines (V_a), total variance (V_t), variance within lines (V_w) and variance between lines (V_b) with random mating in a population of constant size with $q_a = 0.1$.

to 0.5 and then decline. For small values of q, V_a increases as F^2 when F is low. As F approaches unity, V_a becomes equal to three-fourths of V_w. V_b increases slowly at the start as F^3 but increases more rapidly when F is greater than 0.50. V_t increases almost linearly with F in both cases.

Some discrepancies will be noted between the results obtained from the two models. For instance, in the continuous case there is no variation between lines at zero inbreeding whereas there is in the full-sib case. The continuous case, deriving from random mating between a chosen number of parents, automatically includes the possibility of some self-fertilization. The first generation to show variation between lines is therefore inbred. If self-fer-

tilization is explicitly excluded, as of course in a bisexual organism, it is possible to have variation between lines (or rather between families) without the animals being inbred. However, the two systems are otherwise in good agreement.

GENES SHOWING OVERDOMINANCE

It is possible that, at some loci, the phenotypic value of the heterozygote may lie outside the range of those of the homozygotes. Of the extent and kind of such " overdominance " we know very little. As a model, it will be assumed

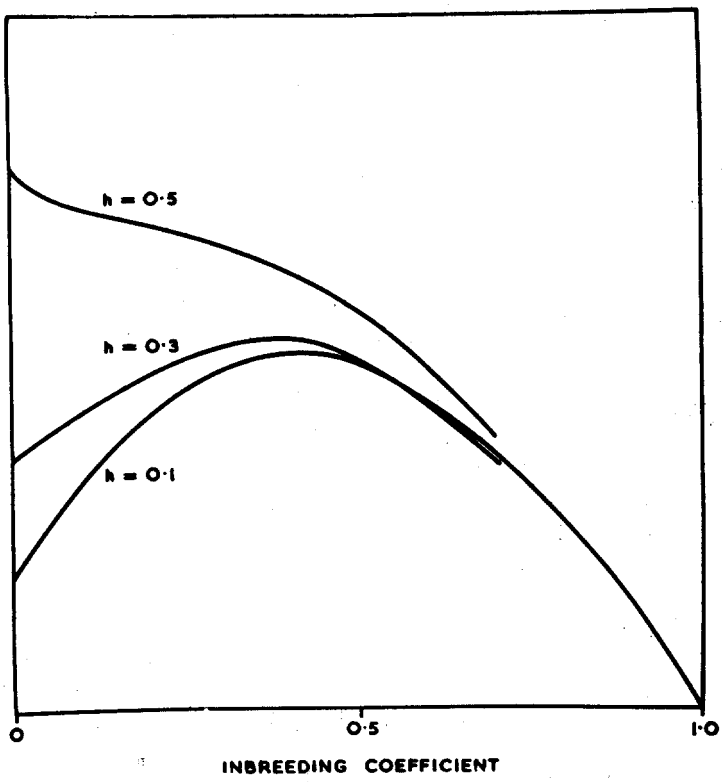

FIGURE 6.—Variance within lines for a locus with overdominance. Phenotypic values are AA = h, Aa = 0, aa = 1 and $q_s = 0.1$.

that the phenotypic values are AA = h, Aa = 0 and aa = 1. Using the moment terminology, the variance between lines is then given by

$$V_w = 2h^2\mu_1 + (1 - 2h - 5h^2)\mu_2 + 4h\,(1 + h)\mu_3 - (1 + h)^2\mu_4.$$

Figures 6 and 7 show the within and between line variances for several values of h when q = 0.10. The curves for h = 0.1 are little different from those for h = 0 in figure 4 but as h increases, the increase of V_w with F becomes less until for a value of h slightly over 0.3, V_w decreases continuously with F. From these results, it seems likely that the general conclusions

arrived at for completely recessive genes will also apply to genes showing overdominance provided h is less than 0.2.

THE EFFECTS OF SELECTION

In the earlier analysis it was assumed that there was no selection against the recessive gene. As, in general, recessives cause some decline in fitness when they are homozygous, it seemed worthwhile to calculate the changes in variance in the extreme case when the selection against the homozygous recessive is complete. Here, the mean frequency of the recessive gene in the

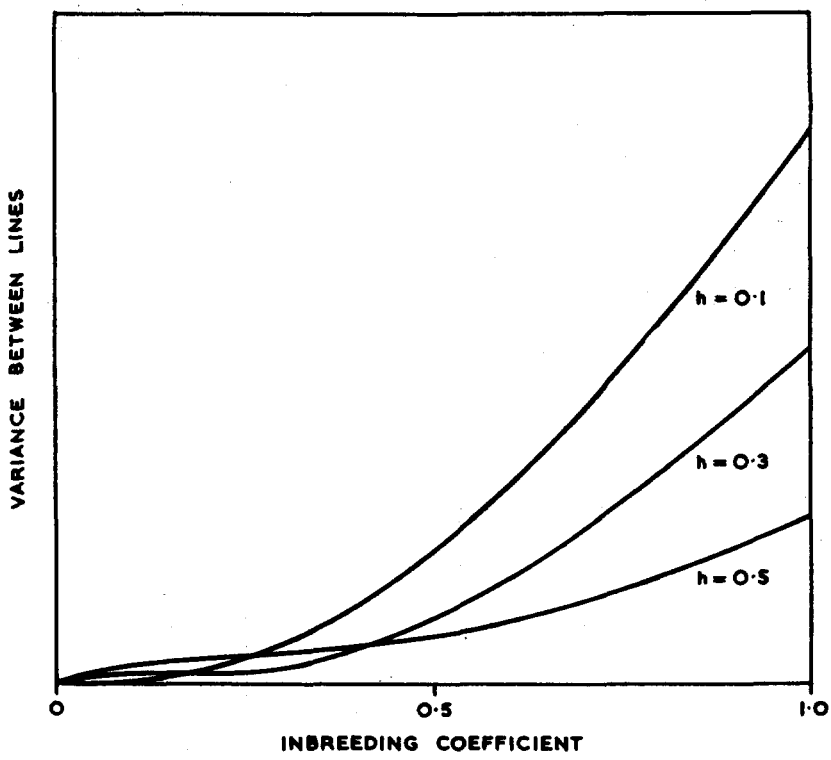

FIGURE 7.—Variance between lines for a locus with overdominance. Phenotypic values are AA = h, Aa = 0, aa = 1 and $q_a = 0.1$.

population of lines will not remain the same but will gradually decline as selection proceeds. The genetic variance will not depend on the inbreeding coefficient alone but also on the amount of selection and therefore on the number of generations that the inbreeding and selection has proceeded. There will thus be no general solution in terms of F and each inbreeding system will have to be treated separately. For continued full-sib matings, when selection against aa animals is on an individual basis, there are only three possible types of mating as shown in table 2.

The average variance within lines is $3u/16$ and that between lines is $u(1-u)/16$. The results obtained are shown in figure 8 for $q = 0.10$ and

TABLE 2

Frequencies of the various mating types with complete selection against the recessive factor.

	AA × AA x_0	AA × Aa y_0	Aa × Aa u_0
x_1	1	1/4	1/9
y_1		1/2	4/9
u_1		1/4	4/9

the curve for V_w from figure 2 is included for comparison. Both V_b and V_w rise at first and then decline to zero as inbreeding approaches completion and all lines become AA in constitution. In the early stages, the effect of selection on the behaviour of the variance within lines is fairly small. V_w rises to 2.69 times its random-breeding value in the first generation and does not decline below the random-breeding value for 7 generations. With no selection, the maximum V_w is 2.99 times the random-breeding value and it takes 10 generations to decline to that value again. The selection against the recessive on this model is the most stringent possible on an individual basis and one can safely make the generalization that in the early generations of full-sib

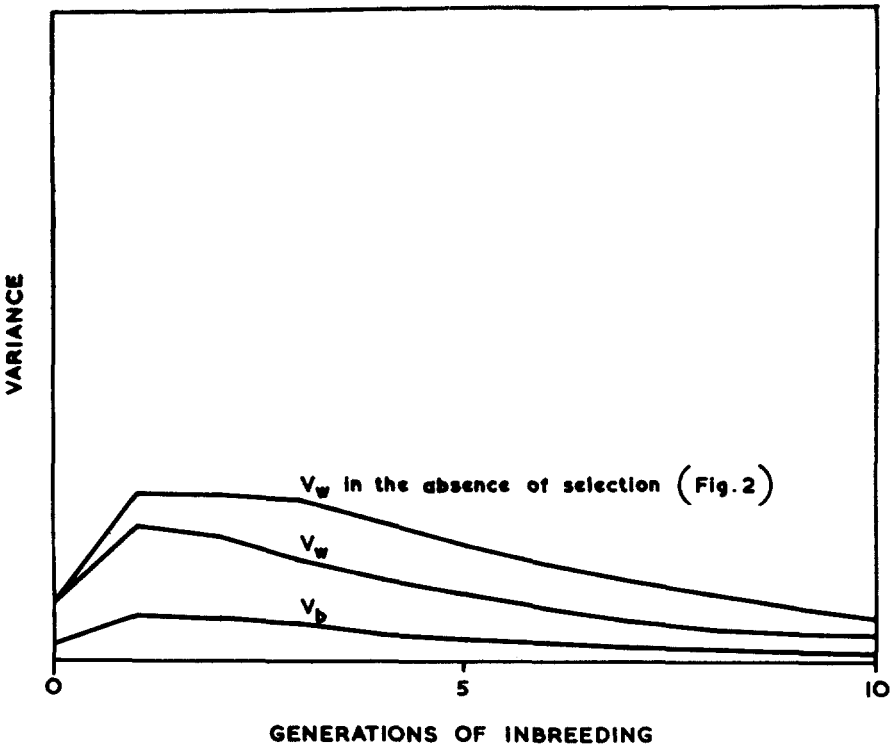

FIGURE 8.—Variance within lines (V_w), and variance between lines (V_b) with full-sib mating and complete selection against the recessive allele, whose initial frequency was 0.1.

mating, selection will not greatly affect the behaviour of the variance within lines. If the inbreeding proceeds more slowly, selection will be more important as it will have more opportunity to take effect.

THE PERFORMANCE OF THE LINES IN CROSSING

From the practical point of view more interest attaches to the performance of the crosses between lines than to that of the lines themselves. The performance of the crosses made between members of a group of lines is often discussed in terms of " general combining ability " and " special combining ability." The " general combining ability " of a line refers to the average performance of the crosses between that line and all the other lines. The " special combining ability " of a particular cross refers to the difference between the performance of the cross and what would have been expected from the general combining abilities of the parent lines. In mathematical terms, the performance of a particular cross P_{ij} between the i^{th} and j^{th} lines is given by

$$P_{ij} = m + a_i + a_j + a_{ij}$$

where m is the mean of all crosses, a_i, a_j are the general combining abilities of the i^{th} and j^{th} lines and a_{ij} is the interaction term, the special combining ability. The term " top-cross " refers to the crosses made between a line and a sample of individuals from the random-bred population. The average top-crossing performance of a line should be equal to the general combining ability in crosses with lines drawn without selection from the random-bred population, because the gametes from a group of inbred lines made without selection are exactly equivalent to a random sample of gametes from the random-bred population. In a similar manner, a series of crosses made at random between completely inbred lines made without selection are equivalent to a group of individuals drawn from the random-bred population.

Consider two lines in which the gene frequencies of the recessive are q_1, q_2. Then, assuming complete dominance, the average performance of the cross between them is $q_1 q_2$. The general combining ability of a line with gene frequency q_1 is $q_1 q$ where q is the average gene frequency in the lines and in the original population. The variance between lines in general combining ability is q^2 var $q_1 = Fq^3(1-q)$. The variance between crosses is var $(q_1 q_2)$ where q_1, q_2 are independent samples from a known distribution. This can be evaluated from the moments about zero of the distribution as $\mu_2^2 - \mu_1^4$ and is equal to $q^2(1-q)F(F+2q-Fq)$, being equal to the variance in the random-bred population, $q^2(1-q^2)$, when F = 1. From the above equation for P_{ij} the variance between crosses due to special combining ability is equal to the total variance between crosses minus twice the variance between lines in general combining ability. This equals $F^2 q^2(1-q)^2$. When F is small compared to q, therefore, the variance between crosses is mostly due to general combining ability but as F increases, the special combining ability becomes much more important. Figure 9 shows the relative contributions that the two parts make to the variance between crosses when $q = 0.10$. From the practical as-

pect, it is the best cross between members of a group of lines that is important. For a given number of lines, the probable superiority of the best cross above the mean will be proportional to the standard deviation between crosses and will be proportional to the first power of F if q is small.

The correlation between the performance of lines in crossing when the lines are partially inbred with the performance when inbreeding is complete is of some practical interest. As inbreeding causes the gene frequencies to deviate between lines but does not change the average gene frequency in the

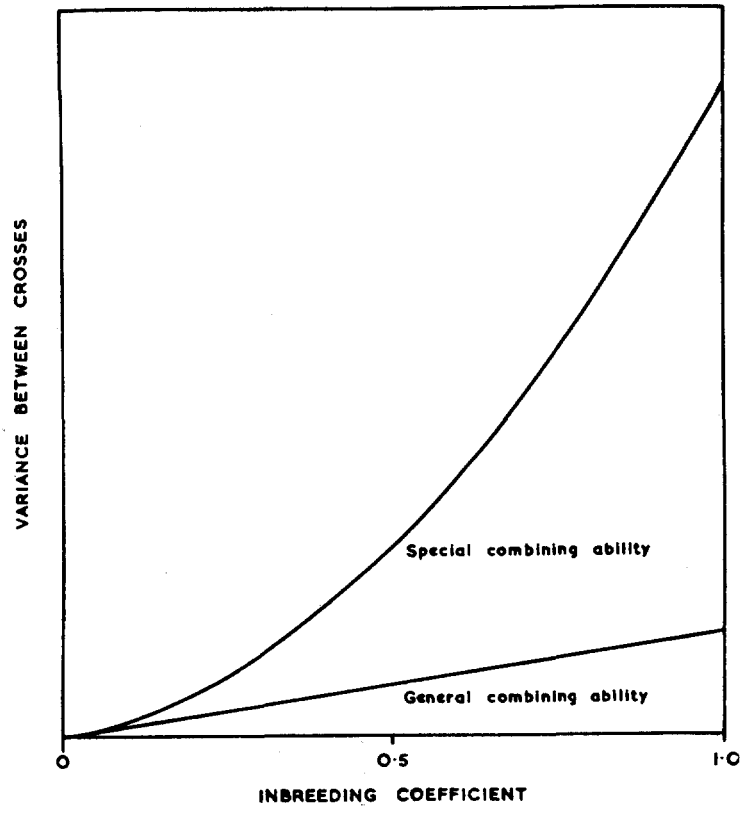

FIGURE 9.—Relative contribution of general and specific combining ability to the variance between crosses with change in inbreeding. $q = 0.1$.

whole population of lines, it follows that the expected gene frequency in completely inbred lines deriving from a given partial inbred line is equal to the gene frequency in that line. Taking the cross between two partially inbred lines in which the frequency of the recessive is q_1 and q_2, we want to know the expected value of $q_1 q_2$ when the inbreeding is complete. Because inbreeding does not change the expected gene frequency, it also does not change the expected value of the cross between two lines in the absence of environmental variation or errors of measurement. In other words, the regression of the

future performance of a cross between two lines on its present performance will always be unity, irrespective of the stage at which the lines are measured. From this, it follows that the correlation between future performance and present performance will be equal to

$$\sqrt{\frac{\text{variance of present performance}}{\text{variance of future performance}}}.$$

In particular, for general combining ability, this equals

$$\sqrt{\frac{Fq^3(1-q)}{q^3(1-q)}} = \sqrt{F}$$

and for the performance of a specific line cross it is

$$\sqrt{\frac{q^2(1-q)\,F(F+2q-Fq)}{q^2(1-q^2)}} = \sqrt{\frac{F(F+2q-Fq)}{1+q}}$$

which is equal to F when q is small.

The correlation between the phenotypic value of a line and its general combining ability will be generally for a single gene fairly close to one, being a correlation between q_1 and q_1^2.

VARIATION DUE TO MANY RECESSIVE GENES

We have been dealing above with the variation due to a single recessive gene. In practice, the genetic variation may be expected to be due to many genes with different gene frequencies and effects of different magnitude. Fortunately, this does not greatly complicate the picture and many of the results can be taken over directly from the single gene case. The resultant variance will be merely the sum of the variance due to the separate genes, so that a generalization can be made about the variation due to recessive genes at frequencies less than about 0.3, that the within line variance will increase until F is in the region of 0.5 and then decline and that the between line variance will increase at first as F^3. Indeed, the presence of variation due to many genes means that as far as the within line variance is concerned, lines will deviate less from the predicted behaviour than they would if the variation were due only to a single gene. In a similar way, the formulae referring to the crossing performance of lines for genes of low frequency can also be taken over to the general case as can those for the correlations and regression of future performance and present performance of crosses.

DISCUSSION

The actual experimental evidence on the effect of inbreeding on the variation within lines is fairly scanty but, in general, the decline in phenotypic variation is slight and in some cases it is known to have increased above the original value after several generations of brother-sister matings (*e.g.*, PEASE

1948). Apart from the possibilities arising from the present paper, there are three other possible causes for such a phenomenon.

(a) Natural selection for heterozygotes may be opposing the trend towards homozygosis produced by inbreeding.

(b) In many characters, the greater part of the variation is environmental in origin and therefore will not be affected by inbreeding. In characters like egg production index in poultry or litter size in swine, the changes in genetic variance may be undetectable against the background of the environmental variance.

(c) The inbred lines may differ from the random-bred stock in their response to environmental changes. WRIGHT (1935) has described a line of guinea-pigs in which a proportion of animals are otocephalic. There is considerable variation in head shape within the line which, on testing, was found to be not genetic in origin. It seems that the line has shifted towards some critical threshold in the process of head formation over which a proportion of the environmental variations takes the animals in the course of development, resulting in a variety of different abnormalities of head shape.

To these three factors affecting the total variation within lines, we may now add a fourth—that the variation due to recessive genes at low frequency will increase with inbreeding until F is about 0.50 and may not return to its original value until F reaches close to 1.

We are still fairly ignorant about the exact behaviour of the genes responsible for continuous variation. In some characters, *e.g.*, fat percentage in milk in cattle, it is likely that the genes are acting mostly in an additive manner. In other, in particular, characters with low heritability that show inbreeding depression, *e.g.*, egg production index in poultry, yield in maize, a high proportion of the genetic variation might be due to recessive or overdominant genes. Such genes will generally be held at a low frequency in the population by natural selection. The possible increase of the genetic variance due to such genes with increasing inbreeding has therefore some practical importance. There are some writers who maintain that animals whose performance is inferior are so because they are homozygous for deleterious recessives. They argue that the only way to improve the general level of the stock is to uncover the recessives by inbreeding and so to produce a population with a uniformly high level of performance. In fact, even with stringent selection against such recessives, it will take several generations of brother-sister matings in which the recessives are segregating out before the genetic variance within such an inbred population will decline to its original value. This is only one of several objections to such a programme.

The results presented here may be of some use in providing a possible explanation for some peculiar experimental results but it is doubtful whether they can be of any precise value in the analysis of continuous variation. When the variation is due to several recessive genes at different frequencies, this treatment can only supply a general description of the probable behaviour of the variances, not of sufficient precision for the experimental results to be used to give much information about the behaviour of the genes themselves.

The only situation in which the gene frequencies are known accurately—in a cross between two inbred lines—the position is complicated by linkage. In discussing the variation due to several genes above, it has been assumed that in the initial random-bred population there is no correlation between the genes present at adjacent loci in a gamete. In the F_2 of a cross between two inbred lines, there will be such a correlation between genes at adjacent loci and any analysis will tell us about the properties of such blocks of genes rather than of the individual genes. In the absence of overdominance at individual loci, such blocks of genes will tend to show overdominance themselves, due to the usual covering-up of recessives. It seems therefore that unfortunately the use of such a cross cannot tell us much about the dominance relationships of the individual genes.

SUMMARY

The effect of inbreeding on the variation due to recessive genes has been treated theoretically both for the case of continued full-sib mating and in lines of small breeding size. If the recessives are at low frequency, the variation within lines increases to a maximum when F is close to 0.50, and declines to zero when inbreeding is complete. The additive component of the variance within lines behaves in a similar manner. The variance between lines is small at first, increasing as F^3 when F is small. The total variance in the population of lines increases almost linearly with F. The variance in the performance of crosses between lines is made up of a component due to the general combining ability of lines proportional to F and to a component ascribable to the special combining ability in particular crosses proportional to F^2. The special combining ability thus becomes much more important as inbreeding progresses. The effects of overdominance and selection are also briefly treated.

LITERATURE CITED

FISHER, R. A., 1949 The Theory of Inbreeding. viii + 120 pp. Edinburgh. Oliver and Boyd.

HALDANE, J. B. S, 1937 Some theoretical results of continued brother-sister mating. J. Genet. **34**: 265–274.

JENNINGS, H. S., 1916 The numerical results of diverse systems of mating. Genetics **1**: 53–89.

PEASE, M. S., 1948 Inbreeding in poultry livestock improvement. Proc. 8th World's Poultry Congress, 33–45.

WRIGHT, S., 1921 Systems of mating. Genetics **6**: 111–178.

WRIGHT, S., and O. N. EATON, 1923 Factors which determine octocephaly in guinea-pigs. J. Agr. Res. **26**: 161–181.

APPENDIX

A GENERAL DERIVATION OF THE RELATIONSHIP OF THE GENETIC VARIANCE TO THE COEFFICIENT OF INBREEDING, F

In terms of gene frequency, the effect of inbreeding can be looked upon as the gradual widening of the distribution of gene frequencies until, when inbreeding is complete, q may only take the values 0 or 1. Starting with a known gene frequency, q, in lines of constant breeding size of n/2 animals, we can consider the second generation as derived from the first by the sampling of groups of n haploid sets, the gene frequency in the different

groups being distributed binomially with mean nq and index n. The next generation is then the repetition of this process, each line giving rise to a group of lines whose gene frequencies are binomially distributed about the mean of the parent line. If the number of lines is constant, the sample of existing lines can be considered as a random sample from the above hypothetical population.

Consider, in the r^{th} generation, the lines (having frequency f_1) in which the gene frequency is q_1. These lines will then by the above operation give a new group of lines in which the moments about zero of the gene frequencies are by the usual formulae for a binomial distribution.

$$[\mu_1]_1 = q_1$$

$$[\mu_2]_1 = \frac{q_1}{n} + \left(1 - \frac{1}{n}\right)q_1^2$$

$$[\mu_3]_1 = \frac{q_1}{n^2} + \frac{3}{n}\left(1 - \frac{1}{n}\right)q_1^2 + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)q_1^3$$

$$[\mu_4]_1 = \frac{q_1}{n^3} + \frac{7}{n^2}\left(1 - \frac{1}{n}\right)q_1^2 + \frac{6}{n}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)q_1^3 + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right)q_1^4$$

The moments of the total population of lines will be the sum of the moments of the groups of lines, arising from each value of q_1 with appropriate weights f_1.

Thus: $$_{r+1}\mu_2 = \sum_q \frac{{}_rf_1\, q_1}{n} + \sum_q \left(1 - \frac{1}{n}\right){}_rf_1\, q_1^2$$

$$= \frac{1}{n}\,{}_r\mu_1 + \left(1 - \frac{1}{n}\right){}_r\mu_2$$ (the r subscripts referring to generations.)

Similarly we have three other equations relating the moments in the $(r+1)^{th}$ generation to those in the r^{th} generation, which can be written diagrammatically as follows:

	$_r\mu_1$	$_r\mu_2$	$_r\mu_3$	$_r\mu_4$
$_{r+1}\mu_1$	1			
$_{r+1}\mu_2$	$\frac{1}{n}$	$1 - \frac{1}{n}$		
$_{r+1}\mu_3$	$\frac{1}{n^2}$	$\frac{3}{n}\left(1 - \frac{1}{n}\right)$	$\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)$	
$_{r+1}\mu_4$	$\frac{1}{n^3}$	$\frac{7}{n^2}\left(1 - \frac{1}{n}\right)$	$\frac{6}{n}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)$	$\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right)$

Thus, knowing the values in the zero generation, we could work out the values in any generation. By the use of matrix theory, it is possible to obtain a general expression for the moments in any generation. For a four-rank matrix such as the above, there are four latent roots, λ_0, λ_1, λ_2, λ_3, and to each latent root there corresponds a latent vector t (a linear function of the moments) such that:

$$_{r+1}t_0 = \lambda_0\, {}_rt_0$$
$$_{r+1}t_1 = \lambda_1\, {}_rt_1$$
$$_{r+1}t_2 = \lambda_2\, {}_rt_2$$
$$_{r+1}t_3 = \lambda_3\, {}_rt_3$$

Knowing the zero value for the latent vectors, t, we can easily calculate the values in the r^{th} generation as ${}_rt_0 = {}_0t_0\,\lambda_0^r$ and so on, and therefore also calculate the moments by expressing them as functions of the t's. As the elements to the right-hand of the diagonal are zero, the four latent roots are simply the entries in the main diagonal. As an example of the evaluation of the latent vectors, consider the vector t_2 corresponding to the root λ_2. We may express t_2 as $a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + a_4\mu_4$. We then obtain the coefficients by equating the coefficients of the μ's deriving from the t recurrence equations with those deriving from the μ recurrence equations. Writing $1 - \frac{1}{n} = \lambda_1$, etc. we have

$$a_4\lambda_2 = a_4\lambda_3$$

$$a_3\lambda_2 = a_4\frac{6\lambda_2}{n} + a_3\lambda_2$$

$$a_2\lambda_2 = a_4\frac{7\lambda_1}{n^2} + a_3\frac{3\lambda_1}{n} + a_2\lambda_1$$

$$a_1\lambda_2 = a_4\frac{\lambda_0}{n^3} + a_3\frac{\lambda_0}{n^2} + a_2\frac{\lambda_0}{n} + a_1\lambda_0$$

giving

$$a_4 = 0$$

$$a_3 = 1 \text{ (an arbitrary value)}$$

$$a_2 = -\frac{3}{2}$$

$$a_1 = \frac{1}{2}$$

Then $$t_2 = \frac{1}{2}\mu_1 - \frac{3}{2}\mu_2 + \mu_3$$

The four equations for the t's are then:

$$\lambda_0 = 1 \qquad t_0 = \mu_1$$

$$\lambda_1 = 1 - \frac{1}{n} \qquad t_1 = -\mu_1 + \mu_2$$

$$\lambda_2 = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \qquad t_2 = \frac{1}{2}\mu_1 - \frac{3}{2}\mu_2 + \mu_3$$

$$\lambda_3 = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right) \qquad t_3 = -\frac{n-1}{5n-6}\mu_1 + \frac{6n-7}{5n-6}\mu_2 - 2\mu_3 + \mu_4$$

If n is large, we may write

$$t_3 = -\frac{1}{5}\mu_1 + \frac{6}{5}\mu_2 - 2\mu_3 + \mu_4$$

Correspondingly, we have four equations for the moments in terms of the latent vectors.

$$\mu_1 = t_0$$

$$\mu_2 = t_0 + t_1$$

$$\mu_3 = t_0 + \frac{3}{2}t_1 + t_2$$

$$\mu_4 = t_0 + \frac{9}{5}t_1 + 2t_2 + t_3$$

In the zero generation, we have ${}_0\mu_n = q^n$ giving

$$
\begin{aligned}
{}_0t_0 &= q \\
{}_0t_1 &= -q(1-q) \\
{}_0t_2 &= q(1-q)\left(\frac{1}{2} - q\right) \\
{}_0t_3 &= -\frac{1}{5}q(1-q) + q^2(1-q)^2
\end{aligned}
$$

Then for V_w in the r^{th} generation,

$$
\begin{aligned}
{}_rV_w &= {}_r\mu_2 - {}_r\mu_4 \\
&= {}_rt_0 + {}_rt_1 - \left({}_rt_0 + \frac{9}{5}{}_rt_1 + 2{}_rt_2 + {}_rt_3\right) \\
&= -\frac{4}{5}{}_rt_1 - 2{}_rt_2 - {}_rt_3 \\
&= -\frac{4}{5}{}_0t_1\,\lambda_1^r - 2{}_0t_2\,\lambda_2^r - {}_0t_3\,\lambda_3^r \\
&= \frac{4}{5}q(1-q)\,\lambda_1^r - q(1-q)(1-2q)\,\lambda_2^r + \left[\frac{1}{5}q(1-q) - q^2(1-q)^2\right]\lambda_3^r
\end{aligned}
$$

Now $\lambda_1^r = \left(1 - \frac{1}{n}\right)^r$ and as $\frac{1}{n}$ is the expected relative decline in heterozygosis each generation, $\left(1 - \frac{1}{n}\right)^r$ is the proportion remaining after r generations and is equal to $1 - F$. If n is large, then $\left(1 - \frac{2}{n}\right)^r = \left(1 - \frac{1}{n}\right)^{2r} = (1-F)^2$. Thus $\lambda_2^r = (1-F)^2$ approximately and $\lambda_3^r = (1-F)^6$, giving

$$V_w = a(1-F) + b(1-F)^3 + c(1-F)^6.$$

Similarly,

$$
\begin{aligned}
{}_rV_a &= 2({}_r\mu_3 - {}_r\mu_4) \\
&= 2\left[{}_rt_0 + \frac{3}{2}{}_rt_1 + {}_rt_2 - \left({}_rt_0 + \frac{9}{5}{}_rt_1 + 2{}_rt_2 + {}_rt_3\right)\right] \\
&= 2\left(-\frac{3}{10}{}_rt_1 - {}_rt_2 - {}_rt_3\right) \\
&= \frac{3}{4}a(1-F) + b(1-F)^3 + 2c(1-F)^6
\end{aligned}
$$

V_t is given by the expression, $V_t = \mu_2 - \mu_2^2$
where $\mu_2 = q - q(1-q)(1-F)$
$= q^2 + q(1-q)F$, giving
$V_t = q(1-q)[q(1+q) + F(1-2q^2) - F^2q(1-q)]$

If q is small, this reduces to Fq. V_b is then obtained as $V_t - V_w$ but unfortunately no simple expression seems to exist. Expansion gives, in order of powers of F,

$$V_b = q(1-q)\left\{F4q^2 + F^2[14q(1-q) - 6q] + F^3[3 + 2q - 20q(1-q)]\right\} + \ldots.$$

If q is small, the first and second terms, being of the third and second order in q, will be small compared to the third term and therefore if F is small, $V_b = 3F^3q$.

For the genes showing overdominance, the general principles are the same except that the values for the first and third moments also enter into the calculation.

The variance between the performances of the lines on top-crossing to the original population can easily be calculated. If the gene frequency in a line is q_1 then the proportion of homozygous recessives in back-crossing to the original population (in which the gene frequency is q) is qq_1, and this will be the mean phenotype value of the cross. The variance required is then $q^2 \text{ var } q_1 = q^2(\mu_2 - \mu_1^2) = q^2[q - q(1-q)(1-F) - q^2] = Fq^2(1-q)$ which increases as the first power of F. If we cross two lines in which the gene frequencies are q_1, q_2, the mean phenotypic value of the cross is q_1q_2. To calculate the total variance between such crosses, we have to find the variance of q_1q_2 when q_1 and q_2 are independent members of the q distribution. Actually the moments about zero of such a distribution of a product are the products of the moments of the parent distributions. As in this case, the two samples are from the same distribution, the moments about zero of the product distribution are the square of the moments of the q distribution. In terms of those moments the variance between line crosses $= \mu_2^2 - \mu_1^4$

$$= q^2(1-q)F(F + 2q - Fq)$$

Thus the variance between line crosses is proportional to F^2 if q is small.