APPLICATION OF GENETICS TO PLANT BREEDING

II. THE INHERITANCE OF QUANTITATIVE CHARACTERS AND PLANT BREEDING[1]

By V. G. PANSE

Galton Laboratory, University College, London

CONTENTS

I. Introduction

It is pointed out in Part I of this contribution and is now generally agreed that the influence of genetics on plant breeding has not fulfilled the early expectations. While there is no doubt that the development of the subject has affected the general outlook of the breeder, the detailed guidance to which he has been looking forward is still lacking. For example, the breeder is now conscious that the selective capacity of his material depends upon the amount of heritable variability present, but it is still necessary to understand in what manner this variability functions, in order to be able to estimate it and use it explicitly in breeding. Similarly, an analysis of the genetic situation in the character under breeding is of vital importance, for it is on such points as the number and magnitude of the factors and their dominance and epistatic relations that the maximum improvement attainable, the rate of improvement by selection and also the most efficient procedure of selection depend. In studying these aspects of the genetical situation and the effect of selection, the influence of the environment cannot be neglected, since, as it is well established,

[1] Read at the Seventh International Genetic Congress, Edinburgh, August 1939.

all quantitative characters are susceptible to the environment and the variability from this source persists in a considerable measure in spite of all the refinement in the experimental technique. A study of the variance in F_2 progenies of crosses between three strains of cotton by Hutchinson *et al.* (1938) can be quoted as an example on this point.

Fisher *et al.* (1932) have pointed out the difficulties in the study of the genetics of quantitative characters and have given an outline of a statistical approach to the problem. They have shown that it is impossible to apply the usual method of genetic analysis here, and, therefore, new and essentially statistical methods must be developed. Not much progress, however, in handling actual experimental data on these lines appears to have been made so far. One of the few examples is "Student's" (1934) estimation of the minimum number of genes determining oil content in Winter's selection experiment. Recently, Charles & Smith (1939) have given statistical criteria for distinguishing between the arithmetic and geometric type of gene action in quantitative inheritance.

The results of a statistical study of the inheritance of staple-length in cotton having a bearing on some of the points mentioned above are reported here. Staple-length is the mean length of the cotton fibres and is the principal character governing spinning quality. The results are by no means complete and are chiefly intended to illustrate a possible mode of attack on the problem of quantitative inheritance. In presenting them the genetical and plant breeding aspects are emphasized and the mathematical details reduced to a minimum. The statistical method used is, however, of a general application and will be published separately.

II. The analysis of experimental data

(*a*) *Experimental material*

The experimental data used in the present work were obtained from crosses between three strains of Indian cotton (*G. arboreum* var. *neglectum*), Bani, Malvi and C 520. The sources of these strains and their agricultural behaviour are described elsewhere (Hutchinson *et al.* 1938). All strains were self-fertilized and line-bred at Indore in Central India for eight generations before the crosses used in the experiments were made in 1934. F_1's were grown in 1935. Taking four F_1 plants from each cross, the resulting F_2 progenies were grown in 1936 in a randomized progeny row trial (Hutchinson & Panse, 1937) with five blocks, each plot consisting of a single row of ten plants. The experiment suffered from defective germination and later from a rather severe attack of wilt, and consequently the number of plants at the time of harvest was only about

50 % of the expected. Ten plants from each F_2 progeny were selected at random, and the F_3 progenies from these were grown in 1937 in two blocks, each progeny-plot consisting of ten plants as before. Instead of randomizing all the 120 progenies together in each block, the twelve families, four from each cross and each family consisting of ten sister progenies arising from a common F_2 progeny, were first randomized and within each family-plot the sister progenies were randomized. The chief object of employing this experimental design was to reduce the environmental contribution to the variances by providing blocks, and further in the F_3 trial, by growing the sister progenies in close proximity of one another. Only self-fertilized seed was used for sowing and seed cotton from selfed flowers was picked for each plant separately and was generally used for examination. The staple-length to which the present data refer was measured on five random seeds per plant. The method of measurement is described elsewhere (Panse, Appendix III to Hutchinson & Ramiah, 1938).

(b) The regression of F_3 progenies on F_2 plants

The analysis of variance of the F_2 data is given below for each cross separately:

	Cross								
	C 520 × Bani			C 520 × Malvi			Malvi × Bani		
Due to ...	D.F.	S.S.	M.S.	D.F.	S.S.	M.S.	D.F.	S.S.	M.S.
Blocks	4	12·844	3·211	4	5·210	1·302	4	5·950	1·487
Progenies	3	0·556	0·185	3	3·213	1·071	3	3·201	1·067
Plot error	12	20·051	1·671	12	13·719	1·143	12	11·943	0·995
Within plots	71	214·085	0·919	101	330·622	0·595	88	212·571	0·526

The first three items were obtained by analysing plot mean values. The sums of squares for the last item were calculated from individual plant values within each plot and pooled over all plots. The corresponding mean squares were further divided by the harmonic means of plant numbers in all plots in order to make them comparable with the mean squares obtained from plot values. A comparison between the mean squares for plot error and "within plots" gives the variance ratios 1·82, 1·92 and 1·89 for the three crosses respectively. The last two are significant on the 5 % level, while the first is slightly lower than the value required for significance. This means that plot to plot differences affect staple-length in addition to differences between plants in the same plot. A similar result was obtained with large plots ($\frac{1}{100}$ acre each) in field scale varietal trials in Central India (Hutchinson & Panse, 1935). In the present case, however, part of the differences is genetic.

The effect of plot differences on staple length is important in relation to the method of selection of the best plants for propagation. The value of the selected plant will partly depend on the plot to which it belongs, and, therefore, the mean value of the plot must also be taken into account. This point was studied in calculating the regression of F_3 progenies on F_2 plants. The relevant portions of the analyses of variance of the F_3 progenies are shown below.

	Cross								
	C 520 × Bani			C 520 × Malvi			Malvi × Bani		
Due to ...	D.F.	S.S.	M.S.	D.F.	S.S.	M.S.	D.F.	S.S.	M.S.
Progenies	35	98·63	2·818	36	106·54	2·959	34	53·38	1·570
within families Plot error	35	21·96	0·628	36	27·71	0·770	34	11·65	0·343

Data of one progeny in the first cross and of two in the third were discarded on account of a doubt regarding their correctness. A portion of the sum of squares between progenies can be accounted for by the regression of the progeny mean on the value of the F_2 parent plant. In calculating this regression the staple-length of the F_2 plant was used as one independent variate and the mean staple-length of the plot to which it belonged as the other, and a partial regression equation was obtained in terms of these two variates. The numerical values of the partial regression coefficients were as follows.

Regression coefficient	C 520 × Bani	C 520 × Malvi	Malvi × Bani
b_1 (plant)	0·5117	0·4816	0·1551
b_2 (plot)	−0·5808	−0·1022	−0·0311

The regression equations can be expressed with a slight modification in a more suitable form. If x_1 is the plant value, x_2 the mean plot value and $z = x_1 - x_2$ then $x_1 = z + x_2$ and,

$$b_1x_1 + b_2x_2 = b_1(z + x_2) + b_2x_2 = b_1(x_1 - x_2) + (b_1 + b_2)x_2.$$

The variate in the first term of the latter expression is now the plant value minus the plot value. The variate in the second term is the plot value as before, but the coefficient of this term is the sum of the two partial regression coefficients. Using the method of solving regression equations given by Fisher (1936, § 29), the standard errors for the regression coefficients in the modified equation are $s\sqrt{c_{11}}$ for b_1 and $s\sqrt{(c_{11} + 2c_{12} + c_{22})}$ for $b_1 + b_2$, where s is the standard deviation obtained from the residual sum of squares between progenies and c_{11}, c_{12} and c_{22}

are the elements of the multiplying matrix. The numerical values of the new coefficients and their standard errors are given below:

Cross	b_1	S.E. of b_1	b_1+b_2	S.E. of b_1+b_2
C 520 × Bani	0·5117	0·1186	−0·0691	0·1267
C 520 × Malvi	0·4816	0·1072	0·3794	0·1680
Malvi × Bani	0·1551	0·1421	0·1239	0·1565

In the first two crosses the coefficient b_1 is highly significant. The coefficient b_1+b_2 is positive and significant in the second cross, while it has a very low negative value in the first. The fact that this latter coefficient has no significant negative value in any of the three crosses, while in one of them it is significantly positive, clearly shows that instead of selecting plants for propagation on the merit of their own values, it is profitable to select them on the basis of the excess of their individual values over the mean values of the plots to which they belong. The plots will not then depress the regression by their negative contribution, while there is a possibility of a positive contribution to the regression from this source through genetic sampling, as the second cross indicates. In the third cross the coefficient b_1 is non-significant and small. One explanation of this result would be that in this cross the phenotypic values of the F_2 plants chosen as progenitors were predominantly influenced by environment and were not, therefore, sufficiently correlated with their genetic values. Why the environmental effect should be so high in this particular cross is not clear.

(c) *Estimation of genetic variance*

The genetic portion of the F_2 variance can be estimated from the regression of F_3 progenies on F_2 parents. If x is the value of the parent plant and y the mean value of the resulting progeny, the regression coefficient of y on x is $\frac{Sy\,(x-\bar{x})}{S\,(x-\bar{x})^2}$. Now x can be considered as made up of a genetic component ξ and a non-genetic modification η, i.e. $x=\xi+\eta$. Then the expected value of y is ξ.

Therefore, the expected value of $y\,(x-\bar{x})=E\,\{\xi\,(\xi-\bar{\xi})+\xi\,(\eta-\bar{\eta})\}$. The term $E\xi\,(\eta-\bar{\eta})=0$, because ξ and η are independent of each other, while $E\xi\,(\xi-\bar{\xi})$ can be shown to be equal to $E\,(\xi-\bar{\xi})^2$, which is the expected genetic variance of the parent plants.

The expected value of $(x-\bar{x})^2=E\,\{(\xi-\bar{\xi})+(\eta-\bar{\eta})\}^2$, which is equal to $E\,(\xi-\bar{\xi})^2+E\,(\eta-\bar{\eta})^2$. The first term represents the genetic and the second non-genetic variance of the parent plants. Therefore $\frac{Sy\,(x-\bar{x})}{S\,(x-\bar{x})^2}$, which is

an estimate of $\frac{E\{y(x-\bar{x})\}}{E(x-\bar{x})^2}$, is also an estimate of $\frac{\text{genetic variance}}{\text{total variance}}$ of the parent plants. When, as in the present case, the parent plants belong to different plots and the regression equation of the type, $y=b_1$ (plant − plot) + (b_1+b_2) plot, considered in the previous section is used, the value of the parent plant can be assumed to contain a genetic component, a non-genetic modification due to variation between plants in the same plot and a second similar modification arising from variation between plots. The expectation of the mean value of the progeny will again be equal to the genetic component of the parent and it can be shown (see Appendix) that the regression coefficients b_1 and b_1+b_2 are expressible in terms of the expected genetic and non-genetic variances of the parents as follows:

$$b_1=\frac{\text{genetic variance}}{\text{total variance within plots}}, \text{ and } b_1+b_2=\frac{\text{genetic variance}}{\text{total variance between plots}}$$

Applied to the present data, the numerical value of the regression coefficient b_1 represents the fraction of the total variance within F_2 plots which is genetic. These are given below. The cross Malvi × Bani is omitted on account of the failure of the regression pointed out before:

Cross	Total variance within plots, per plant	b_1	Genetic variance of F_2	Non-genetic variance within F_2 plots
C 520 × Bani	3·015	0·5117	1·543	1·472
C 520 × Malvi	3·273	0·4816	1·576	1·697

Following Fisher (1930, p. 33), the term "genetic" is used here with a special significance and refers to the effects of a strictly additive action of the genes concerned. It is in this sense that the mean value of the progeny is equal to the genetic value of the parent. The genetic variance of the parents estimated on this basis is also the result of the additive action of the genes and is distinguished from "genotypic" variance which is the total variance due to the segregation of genes and includes, in addition to the genetic variance, variance arising from non-additive interactions of genes such as dominance and epistacy. The presence of these interactions tends to depress the regression of the progeny mean on the parental value by causing the genotypic variance of the parents to be greater than the genetic.

The non-genetic fraction of the F_2 variance shown above may thus include the difference between the genotypic and genetic variance in addition to the environmental variance. The fact that it is as high as half of the total variance is important, since selection of the desired genetic values is rendered more and more uncertain as the non-genetic com-

ponent of the variance increases in proportion to the genetic, and consequently the rate of progress by selection is lowered.

(d) *Effective number of factors*

The mean genotypic variance within F_3 progenies is half of the F_2 genotypic variance (Fisher *et al.* 1932), while the variance of the mean variance within F_3 can be shown to be equal to one-quarter of the sum of squares of the variances due to individual factors. If, for example, the variances due to two factors **A** and **B** are **a** and **b** respectively, the variance within F_2 progenies will be due to the segregation of the heterozygotes only, progenies of homozygous F_2 individuals having no variance. Now,

F_3 progenies heterozygous for	Frequency of the progenies	Variance within the progeny
A and **B**	$\frac{1}{4}$	$a+b$
A only	$\frac{1}{4}$	a
B only	$\frac{1}{4}$	b
Neither	$\frac{1}{4}$	0
Mean variance within F_3 progenies		$\frac{1}{2}(a+b)$

And the variance of this variance =

$$\tfrac{1}{4}(a+b)^2+\tfrac{1}{4}a^2+\tfrac{1}{4}b^2-\{\tfrac{1}{4}(a+b)+\tfrac{1}{4}a+\tfrac{1}{4}b\}^2=\tfrac{1}{4}(a^2+b^2).$$

If the two factors are linked a different result is obtained; because now,

F_3 progenies heterozygous for	Frequency of the progenies	Variance within the progeny
A and **B**	$\frac{1}{2}(p^2+q^2)$	$a+b$
A only	pq	a
B only	pq	b
Neither	$\frac{1}{2}(p^2+q^2)$	0

where p and q are the fractions of gametes in coupling and repulsion respectively. While the mean variance within F_3 is $\frac{1}{2}(a+b)$ as with unlinked factors, the variance of this variance is

$$\tfrac{1}{2}(p^2+q^2)(a+b)^2+pqa^2+pqb^2-\{\tfrac{1}{2}(p^2+q^2)(a+b)+pqa+pqb\}^2,$$

which can be shown equal to,

$$\tfrac{1}{4}(a^2+b^2)+\tfrac{1}{2}(p-q)^2\,ab.$$

With linkage, therefore, the variance of the F_3 variance is increased by a quantity $\frac{1}{2}(p-q)^2\,ab$. Below are given the values of this increase as a fraction of ab for various cross-over fractions:

q	$\frac{1}{2}(p-q)^2$
0	0·5
0·1	0·32
0·2	0·18
0·3	0·08
0·4	0·02
0·5	0

With close linkage, i.e. up to 20 % crossing-over, the increase in variance is appreciable but is insignificant beyond this point. With close linkage, however, the two factors will behave in segregation as a single factor and can be effectively counted as one. Therefore, the increase in variance due to linkage need not be taken into consideration in estimating the number of independent factors concerned with a given character.

As shown above, if the F_2 variance is $(a+b+\ldots)$, the mean variance within F_3 progenies is $\frac{1}{2}(a+b+\ldots)=V_3$, and the variance of this variance is $\frac{1}{4}(a^2+b^2+\ldots)=V(V_3)$. Assuming the segregation in F_2 to be due to n factors all with equal variance α, then, $V_3=\frac{n\alpha}{2}$ and $V(V_3)=\frac{n\alpha^2}{4}$.

$$\therefore\ V^2_3/V(V_3)=\frac{n^2}{4}\alpha^2\Big/\frac{n}{4}\alpha^2=n,$$

where n the number of factors, hypothetically with equal variance and without linkage, can be termed the "effective" number of factors. This number can thus be calculated if the mean genotypic variance within F_3 progenies and the variance of the mean variance within F_3 progenies is known. These quantities are shown below for the staple-length data assuming the genotypic variance to be the same as genetic.

Cross	V_3 (= ½ genotypic variance in F_2)	$V(V_3)$	$V^2_3/V(V_3)=n$
C 520 × Bani	0·772	0·363	1·64
C 520 × Malvi	0·788	0·224	2·77

This calculation does not mean that the actual number of factors is necessarily only two or three in the above crosses. The effective number of factors n is merely a ratio of the square of the mean genotypic variance within F_3 to its variance and is, therefore, equivalent to any number of factors with different magnitudes that satisfy this ratio. Another condition governing such a series of factors is obviously a given F_2 genotypic variance. Using the C 520 × Malvi cross to supply the data for these conditions, alternative possibilities in respect to the genetic constitution of the F_2 which satisfy them will be considered in the next section. For simplification of calculations the genotypic variance in F_2 will be taken as 1·5 and an environmental component of equal magnitude. The value of n will be equated to 3, the nearest whole number to the calculated ratio.

III. Models of genetic factors

In setting up hypothetical systems of factors for staple-length segregation in the C 520 × Malvi cross, the two extreme cases considered are (1) three equal factors, i.e. the actual number of factors equal to the

effective number, and (2) a very large number of unequal factors amounting to an infinite series. Assuming the F_2 distribution symmetrical, this can be due either to the absence of dominance in the factors concerned or due to equal dominance in opposite directions, the two types of dominance balancing each other. The latter case deserves special notice for, as will be seen later, it really makes possible the consideration of consequences of a dominance bias in one direction and thus, while keeping the algebraic calculations much simpler, makes a separate examination of the cases with dominance in one direction only, unnecessary in principle. The infinite series of factors will be represented by a single convergent geometric series without dominance and by two such equal series with dominance in opposite directions.

In forming a geometric series of factors with F_2 variance equal to $\frac{3}{2}$ and $V^2_3/V(V_3)=3$, if $1/r$ is the common ratio between the variances of individual factors and a the variance of the first member of the series, it can be shown that $\frac{r+1}{r-1}=V^2_3/V(V_3)=3$, and the F_2 variance $=\frac{ar}{r-1}=\frac{3}{2}$. Therefore, $1/r=\frac{1}{2}$ and $a=\frac{3}{4}$. The factors without dominance giving these variances are, therefore, arranged in a geometric series beginning with $\sqrt{\frac{3}{2}}$ and having the common ratio $\sqrt{\frac{1}{2}}$. With two equal geometric series the F_2 variance due to each is $\frac{3}{4}$ and the ratio $\frac{r+1}{r-1}=\frac{3}{2}$, making the first variance in each series $\frac{3}{5}$ with a common ratio $\frac{1}{5}$. With complete dominance in opposite directions factors in each series, therefore, begin with a magnitude of $2/\sqrt{5}$ and have the common ratio $1/\sqrt{5}$.

The various systems of factors considered are,

I. No dominance, factors in a geometric series and represented as,

AA	$+\sqrt{\frac{3}{2}}$	**BB**	$+\sqrt{\frac{3}{4}}$	**CC**	$+\sqrt{\frac{3}{8}}$
Aa	0	**Bb**	0	**Cc**	0
aa	$-\sqrt{\frac{3}{2}}$	**bb**	$-\sqrt{\frac{3}{4}}$	**cc**	$-\sqrt{\frac{3}{8}}$

etc., with the common ratio $1/r=1/\sqrt{2}$.

The highest possible genotypic value on this set up will be,

$$\frac{\sqrt{\frac{3}{2}}}{1-1/\sqrt{2}}=\frac{\sqrt{3}\,(\sqrt{2}+1)}{(\sqrt{2}-1)\,(\sqrt{2}+1)}=4{\cdot}182,$$

corresponding to the genotype **AABBCC**....

II. No dominance, three equal factors which of course have equal variances. The factors are

AA	$+1$	**BB**	$+1$	**CC**	$+1$
Aa	0	**Bb**	0	**Cc**	0
aa	-1	**bb**	-1	**cc**	-1

the highest genotype **AABBCC** having the value 3.

III. Balanced dominance, factors in two equal geometric series with complete dominance in opposite directions. Each series will contribute $\frac{3}{4}$ to F_2 variance and have the ratio $n=\frac{3}{2}$. The two series can, therefore, be represented as follows.

$$\left.\begin{array}{llll} \text{Series I:} & \mathbf{AA}+2/\sqrt{5} & \mathbf{BB}+\frac{2}{5} & \mathbf{CC}+2/5\sqrt{5} \\ & \mathbf{Aa}+2/\sqrt{5} & \mathbf{Bb}+\frac{2}{5} & \mathbf{Cc}+2/5\sqrt{5} \\ & \mathbf{aa}-2/\sqrt{5} & \mathbf{bb}-\frac{2}{5} & \mathbf{cc}-2/5\sqrt{5} \end{array}\right\} \text{etc.,}$$

$$\left.\begin{array}{llll} \text{Series II:} & \mathbf{A'A'}+2/\sqrt{5} & \mathbf{B'B'}+\frac{2}{5} & \mathbf{C'C'}+2/5\sqrt{5} \\ & \mathbf{A'a'}-2/\sqrt{5} & \mathbf{B'b'}-\frac{2}{5} & \mathbf{C'c'}-2/5\sqrt{5} \\ & \mathbf{a'a'}-2/\sqrt{5} & \mathbf{b'b'}-\frac{2}{5} & \mathbf{c'c'}-2/5\sqrt{5} \end{array}\right\} \text{etc.,}$$

each with common ratio $1/r=1/\sqrt{5}$. The highest genotype

$$\mathbf{AAA'A'BBB'B'CCC'C'}\ldots$$

will be scored as,

$$2\times\frac{2/\sqrt{5}}{1-1/\sqrt{5}}=\sqrt{5}+1=3{\cdot}236.$$

IV. Balanced dominance, three factors with equal variance, one of them without dominance and the other two with complete dominance in opposite directions. They are,

$$\begin{array}{llll} \mathbf{AA}+\sqrt{\frac{2}{3}} & \mathbf{BB}+1 & \mathbf{CC}+\sqrt{\frac{2}{3}} \\ \mathbf{Aa}+\sqrt{\frac{2}{3}} & \mathbf{Bb}\quad 0 & \mathbf{Cc}-\sqrt{\frac{2}{3}} \\ \mathbf{aa}-\sqrt{\frac{2}{3}} & \mathbf{bb}-1 & \mathbf{cc}-\sqrt{\frac{2}{3}} \end{array}$$

The highest genotype **AABBCC** has the value 2·633.

V. Balanced dominance, three factors of equal magnitude with dominance as in the last case. These are,

$$\begin{array}{llll} \mathbf{AA}+\sqrt{\frac{3}{4}} & \mathbf{BB}+\sqrt{\frac{3}{4}} & \mathbf{CC}+\sqrt{\frac{3}{4}} \\ \mathbf{Aa}+\sqrt{\frac{3}{4}} & \mathbf{Bb}\quad 0 & \mathbf{Cc}-\sqrt{\frac{3}{4}} \\ \mathbf{aa}-\sqrt{\frac{3}{4}} & \mathbf{bb}-\sqrt{\frac{3}{4}} & \mathbf{cc}-\sqrt{\frac{3}{4}} \end{array}$$

This case which is a variation of case IV is also comparable to case II on account of the equal size of the three factors; but it should be noted that the ratio n is here 32/11 which is very slightly less than the required value. The highest genotypic value is 2·598.

The statistical consequences in the F_3 generation of selecting as progenitors a certain proportion of individuals in a given range of the F_2 populations with these hypothetical genetic constitutions can be studied with the help of a moment generating function. The F_3 properties determined are (1) the mean value of the progenies, (2) the mean genotypic

variance within the progenies, (3) the covariance of means and variances of the progenies, (4) the variance of the mean values of the progenies, and (5) the variance of the mean variance within the progenies. A selection of 10 % of the F_2 individuals with the highest phenotypic values was assumed for this purpose.

It is not intended to enter into any detail of the calculation here, but the method is briefly indicated. For each of the genetic models an exact simultaneous distribution of the F_2 phenotypic values, the mean values of the resulting progenies and the genotypic variance within these progenies can be set down in the form of a moment-generating function.

TABLE I

Highest genotypic values and the statistical properties of F_3 from selected F_2's with different genetic constitutions

	System of factors	Highest genotypic value	Genotypic mean of F_3 progenies	Mean genotypic variance within F_3 progenies	Variance of the genotypic mean of F_3 progenies	Covariance between F_3 means and variances	Variance of the mean genotypic variance within F_3 progenies
I.	No dominance, factors in a geometric series	4·182	1·512	0·595	0·744	−0·182	0·160
II.	No dominance, three equal factors	3·000	1·512	0·598	0·749	−0·196	0·158
III.	Balanced dominance, factors in two geometric series	3·236	1·266	0·642	0·669	−0·125	0·146
IV.	Balanced dominance, three factors with equal variance	2·633	1·347	0·629	0·693	−0·155	0·153
V.	Balanced dominance, three equal factors	2·598	1·328	0·632	0·697	−0·146	0·153

By expanding it, it is possible to obtain the various sums of powers and of products necessary to express the F_3 properties or quantities related to them in terms of the F_2 phenotypic values by means of a regression equation of the type, $Y = A + Bx + Cx^2 + Dx^3 + Ex^4 \ldots$, where Y represents the particular F_3 quantity and x the F_2 phenotypic value. An equation of the fourth degree was found satisfactory. Since Y is expressed in terms of x whose exact distribution is known, the mean value of Y over a particular range of this distribution can be determined by integrating between the limits appropriate to the degree of selection applied. The assumption of a symmetrical F_2 distribution has made the algebraic work much easier but such an assumption is not necessary to use the method. The results obtained are given in Table I.

These theoretical or population values can be used for two purposes. The last five quantities can be easily calculated from experimental

data on F_2 and F_3 progenies, the latter grown from 10 % of the F_2 individuals selected for highest phenotypic values. By a comparison of the two sets of values it should be possible to discover the presence or absence of dominance in the experimental material, and whether the observed segregation is due to a few large factors or a large number of factors of varying magnitudes. Secondly, the values set out in the table can be used to compare the effects on F_3 of applying selection to F_2 populations of different genetic constitutions and to obtain some indication of the course of further selection.

For comparing experimental with theoretical values, the standard error can be used for the mean value of the F_3 progenies and a χ^2 test for the three variances, since the sum of squares obtained from n sample values divided by the population variance is a χ^2 with $n-1$ degrees of freedom. The covariance cannot be compared directly because it is dependent on the variance of mean values and of variances of F_3 progenies. Where both these variances agree with their theoretical values, the covariance can be tested as a correlation coefficient and where only one of them agrees as a regression coefficient. A separate test of covariance is of no use when both variances differ significantly from their population values. In the present case a comparison of the three variances with their experimental estimates is not likely to be helpful as in these quantities the five models differ very little from one another. The mean value of F_3 progenies and the covariance on the other hand show somewhat wider differences and appear more useful. The experimental data to make the comparisons are not available.

The highest genotypic values given in the table show the limits to which selection can be carried in each case. For equivalent number of factors these limits are higher in the absence of dominance than in its presence, and within each type higher when a given genotypic variance is caused by a large number of factors with varying magnitudes than when it is due to a few large factors. The genotypic mean values of the F_3 progenies show the advance resulting from selection as the mean value of the whole F_3 is zero in each case. A greater progress has been made in the absence of dominance than in its presence, and in the latter case three factors have raised the mean value higher than an infinite series. Considered in relation to the highest values attainable, the progress made in F_3 with three factors with and without dominance represents 51 and 50 % of the maximum respectively, and 39 and 36 % with an infinite series of factors. The ratio $V^2_3/V\,(V_3)$ calculated from the table indicates on the same scale as in the whole F_3 population, the "effective" number of

factors operating in the selected portion of the population. Its calculated values, which are given below, show the greatest reduction in heterozygosity when dominance is absent, and in its presence a greater reduction with three factors than with an infinite number.

System of factors	V^2_3/V (V_3)
I. No dominance, geometric series	2·210
II. No dominance, three equal factors	2·265
III. Balanced dominance, geometric series	2·835
IV. Balanced dominance, three factors with equal variances	2·586
V. Balanced dominance, three equal factors	2·611

The genotypic variance within the F_3 progenies is now approximately of the same magnitude as that between progenies, whereas in the whole F_3 the latter is twice as great as the former in the absence of dominance and one and a half times as great in its presence (Fisher *et al.* 1932). Even so, further selection based on progeny means will be more efficient than selecting individual values, because the latter will be affected by environmental variation to a considerably greater extent. The negative covariance between progeny means and variances in all the five cases indicates on average a falling off of the genotypic variance of a progeny as its mean value increases. To achieve its object selection of progenies with high mean values is necessary, but this at the same time enforces selection of individuals in progenies with a relatively low genotypic variance, and a larger number of individuals must, therefore, be selected to ensure high genotypic values. The weight to be attached to the progeny mean and the individual plants selected for further propagation must be studied and a balance maintained between the two.

The present analysis must be extended to F_4 and subsequent generations to study the rate at which the different properties of the populations with different genetic constitutions are modified under selection. It is also possible that the various statistical quantities calculated for F_3 will show a wider distinction between the different genetic systems at some other level of selection in F_2, than the one adopted.

IV. Discussion

In Part I, Hutchinson has discussed the superiority of progeny row breeding to mass selection and the advantage of selecting on the basis of the progeny means. When selection is to be continued in the progeny, the usual practice is to select a certain number of individuals with the desired phenotypic values and grow further progenies from them. On account of the amount of material involved it is usually necessary to restrict the number of individuals whose progenies can be carried forward, and hence

it becomes all the more important that the individuals possess a high genetic potentiality. The result obtained with the regression of F_3 progenies on F_2 plants shows that when a progeny is grown in replicated plots, selection of individuals in the progeny should be based on the excess of the individual value over the mean value of the plot to which it belongs, as by so doing, the environmental influence on individual values of differences between plots can be eliminated.

In the present study the different Mendelian factors are supposed to act independently and the complicating feature of interaction between factors has been excluded. The phenomenon of dominance or the interaction between the phases of the same factor is, however, taken into account. Fisher (1918) has pointed out that the statistical effects of dominance and epistacy are similar. He has shown that the hypothesis of cumulative Mendelian factors fits the inheritance of human stature very accurately and it is important to examine how far it can explain the observed facts in plant breeding. The effects of dominance on selection have been clearly brought out in the present results. Before discussing them it should be noted in what manner dominance plays its part in the genetic models with balanced dominance. In the F_2 distribution heterozygotes with dominance for low values will accumulate in the lower portion and those with dominance for high values in the upper part. Selection for high phenotypic values will, therefore, almost entirely be restricted to phenotypes with dominance for high values and this dominance will show its effects in the F_3. Thus the assumption of a symmetrical F_2 distribution has not prevented the consideration of dominance.

The table of F_3 values shows the retarding effect of dominance on selection. Comparing the two groups of cases with and without dominance, it will be seen that while starting with the same amount of genotypic variance in the F_2, selection has raised the mean value of the F_3 higher in the absence of dominance than in its presence. In the former case, however, the whole of the F_2 genotypic variance is genetic while in the latter the genetic variance is lower, a part of the genotypic variance being due to dominance. This demonstrates how it is the genetic portion of the variance that determines the immediate capacity of the material for selection. The lower value of the covariance between means and variances of F_3 progenies in the presence of dominance also suggests that a part of the genotypic variance is not associated with mean values. Dominance also counteracts the other effect of selection viz. increase of homozygosity. It will be seen that a greater amount of genotypic variance persists in the F_3 when dominance is present and the ratio

V^2_3/V (V_3) is similarly higher. This ratio, it will be remembered, corresponds to the effective number of factors. Dominance thus appears to slow down the change both in mean values and variances brought about by selection.

The estimate of genetic variance made by the regression of F_3 progenies on F_2 plants was also used as the genotypic variance in studying the different genetic systems. This was necessary in the absence of suitable experimental data for estimation of the latter. Actually, as pointed out before, the genotypic variance will be somewhat greater than genetic in the presence of dominance and has, therefore, been underestimated. In the three cases with balanced dominance it can be shown that the genetic variance of the F_2 is 22–33 % lower than the genotypic. The discrepancy between the two will become smaller if dominance is partial instead of complete as assumed here. For an accurate estimation of genotypic variation in F_2 it is necessary to grow progenies resulting from crossing F_2 plants among themselves at random. Then the F_2 genotypic variance is equal to twice the difference between the covariance of F_2 parental value with the mean of the F_3 offspring and the covariance of the F_2 parental value with the mean of its biparental offspring (Fisher *et al.* 1932).

On the question of the estimation of genotypic variance the paper by Charles & Smith, referred to in the Introduction, is of some interest. In developing criteria for distinguishing between the arithmetic and geometric types of gene action they are led to estimate the genotypic variance by taking the difference between the total and the environmental variance. They calculate the latter on the assumption that it is correlated with mean values in the three non-segregating generations viz. the two parents and the F_1, and the same correlation holds in the segregating generations. In the absence of sufficient experimental evidence to justify this assumption their tests involving variances and skewness are of doubtful value. Under normal field-experimental conditions mean values and variances do not appear to show any such systematic relationship and the two are, therefore, commonly treated as independent. In the staple-length data of the cotton crosses no consistent relationship was observed between them in the parental and F_1 generations tested simultaneously in replicated trials for two years.

The hypothesis of geometrical effects used by these authors is that a given gene substitution multiplies the phenotypic value by a constant quantity characteristic of each factor. On this hypothesis they show that the mean value of the F_1 should be equal to the geometric mean of the two

parental values. As the geometric mean is always smaller than the corresponding arithmetic mean, this hypothesis will not be applicable to the large number of characters in which the F_1 has a value equal to or generally greater than the arithmetic mean of the two parents. Their tests involving mean values for the arithmetic effects of genes are capable of a general application with any type and magnitude of dominance. On the arithmetic hypothesis the mean value of the F_2 should be equal to the mean of the F_1 and the parental values, i.e. $F_2 = \frac{P_1 + P_2 + 2F_1}{4}$. Similarly, the two back cross means should each be equal to the mean of the F_1 and the parent involved. Using these tests on the staple-length data in the three crosses with F_2 values for two years and backcross values for one year, the agreement between the actual and expected values is found very close, the mean difference between the two being, $0{\cdot}087 \pm 0{\cdot}175$ and $0{\cdot}173 \pm 0{\cdot}202$, respectively.

V. Summary

The importance of the study of quantitative inheritance for a closer application of genetics to plant breeding has been recognized. The object of the present paper is to summarize the results obtained in a statistical study of quantitative inheritance relating to F_2 and F_3 progenies. The genetic and plant breeding aspects of the results are emphasized.

The experimental data used refer to the staple-length measurements on F_2 and F_3 progenies of crosses between strains of cotton belonging to the species *G. arboreum* var. *neglectum*, grown at the Institute of Plant Industry, Indore, Central India. The regression of mean staple-length of F_3 progenies on F_2 phenotypic values shows that it is advantageous to consider plot values and select individuals on the basis of their excess over the former where, as in the present case, inter-plot variation affects the character in addition to intra-plot variation. The coefficient of regression also gives an estimate of the genetic fraction of the total F_2 variance. This is an important relationship, as it affords a basis for separating the inheritable and non-inheritable components of variance in the experimental material.

The ratio of the square of the genotypic variance within F_3 progenies to the variance of this variance is shown to represent the "effective" number of factors which can account for the segregation in F_2 and which hypothetically possess equal variance and are without linkage. With a given F_2 variance and a given effective number of factors, it is possible to set up different genetic systems or models consisting of factors varying

in magnitude and number and with or without dominance. Five models with the smallest possible number of factors or an infinite number and each with or without dominance are considered.

By the use of a moment-generating function in three variables the moments of the distribution of the F_2 phenotype and of certain related quantities are calculated for each system. With these it is possible to express in terms of the F_2 phenotype properties of the F_3 progenies such as the genotypic mean or variance, and further to calculate their mean values in a portion of the F_3 population resulting from a selected proportion of F_2 phenotypes. Similar mean values obtained from experimental data can be compared with these theoretical values in order to discover the presence or absence of dominance and to decide on the possible number of factors operating in the experimental material.

In the present case, assuming a 10 % selection in F_2, theoretical mean values for (1) the genotypic mean and (2) variance of F_3 progenies, (3) the covariance between the two, and (4) the variance of F_3 means and (5) of F_3 variances are calculated for each model. Their usefulness for identifying the genetic situation and the information obtained from them on the effect of selection applied and on questions relating to further selection are discussed.

The results summarized in the present paper form part of the work being carried out under the guidance of Prof. Fisher to whom I am indebted for much help. I should also like to thank my colleague Mr D. J. Finney at the Galton Laboratory for his help in discussing with me several mathematical points arising in the course of the work.

REFERENCES

CHARLES, D. R. & SMITH, H. H. (1939). *Genetics*, **24**, 34.

FISHER, R. A. (1918). *Trans. Roy. Soc. Edinb.* **52**, 399.

—— (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

—— (1936). *Statistical Methods for Research Workers*, 6th ed. London: Oliver and Boyd.

FISHER, R. A., IMMER, F. R. & TEDIN, OLOF (1932). *Genetics*, **17**, 107.

HUTCHINSON, J. B. & PANSE, V. G. (1935). *Ind. J. agric. Sci.* **5**, 545.

—— —— (1937). *Ind. J. agric. Sci.* **7**, 531.

HUTCHINSON, J. B., PANSE, V. G. & GOVANDE, G. K. (1938). *Ind. J. agric. Sci.* **8**, 757.

PANSE, V. G. (1938). Appendix III to Hutchinson, J. B. and Ramiah, K., *Ind. J. agric. Sci.* **8**, 582.

"STUDENT" (1934). *Ann. Eugen., Lond.*, **6**, 77.

APPENDIX

To obtain the values of the partial regression coefficients b_1 and b_2, it is necessary to solve the following simultaneous equations,

$$b_1 Sx_1^2 + b_2 Sx_1 x_2 = Sx_1 y, \qquad \ldots\ldots \text{(I)}$$

$$b_1 Sx_1 x_2 + b_2 Sx_2^2 = Sx_2 y, \qquad \ldots\ldots \text{(II)}$$

where y is the mean value of the progeny, x_1 the value of the parent plant and x_2 the mean value of the plot to which it belongs.

The expectations of sums of squares and of products in these equations are considered below.

Let there be n plants in each of m plots. If $\bar{x}$ is the mean value of all mn plants, the value of any plant (x_1) can be scored as $\bar{x}+g+a+b$ and the mean value of a plot (x_2) as $\bar{x}+\frac{1}{n}(Sg+Sa+nb)$, g representing the genetic and a and b the environmental modifications due to variation within and between plots respectively. The corresponding expected variances can be set down as,

$$V_g = \frac{1}{mn} SSg^2, \quad V_a = \frac{1}{mn} SSa^2, \quad V_b = \frac{1}{mn} Snb^2 = \frac{1}{m} Sb^2.$$

These being independent of one another, the total variance between plants is $V_g + V_a + V_b$.

The covariance between plants and plot mean values

$$= \frac{1}{mn} SS\{(x_1 - \bar{x})(x_2 - \bar{x})\}$$

$$= \frac{1}{mn^2} SS\{(g+a+b)(Sg+Sa+nb)\}$$

$$= \frac{1}{mn^2} S\{(Sg+Sa+nb)(Sg+Sa+nb)\}$$

$$= \frac{1}{mn^2} (SSg^2 + SSa^2 + n^2 Sb^2)$$

$$= \frac{1}{mn^2} (mn V_g + mn V_a + mn^2 V_b)$$

$$= \frac{1}{n}(V_g + V_a) + V_b.$$

The variance between plot mean values

$$= \frac{1}{mn} . nS\,(x_2 - \bar{x})^2$$
$$= \frac{1}{m} S \left\{\frac{1}{n}\,(Sg + Sa + nb)\right\}^2$$
$$= \frac{1}{mn^2}\,(SSg^2 + SSa^2 + n^2 Sb^2)$$
$$= \frac{1}{n}\,(V_g + V_a) + V_b.$$

The covariance between parent plants and progeny means

$$= \frac{1}{mn}\,SS\,\{(x_1 - \bar{x})\,(y - \bar{y})\}$$
$$= \frac{1}{mn}\,SS\,\{(g + a + b)\,g\},$$

since the mean value of the progeny is equal to the genetic component of the parent plant,

$$= \frac{1}{mn}\,SSg^2$$
$$= V_g.$$

The covariance between plot and progeny means

$$= \frac{1}{mn}\,SS\,\{(x_2 - \bar{x})\,(y - \bar{y})\}$$
$$= \frac{1}{mn^2}\,SS\,\{(Sg + Sa + nb)\,g\}$$
$$= \frac{1}{mn^2}\,SSg^2$$
$$= \frac{1}{n}\,V_g.$$

Substituting these expectations in equations (I) and (II) we get,

$$(V_g + V_a + V_b)\,b_1 + \left\{\frac{1}{n}\,(V_g + V_a) + V_b\right\} b_2 = V_g, \quad \ldots\ldots(III)$$

$$\left\{\frac{1}{n}\,(V_g + V_a) + V_b\right\} b_1 + \left\{\frac{1}{n}\,(V_g + V_a) + V_b\right\} b_2 = \frac{1}{n}\,V_g. \quad \ldots(IV)$$

Subtracting (IV) from (III) and multiplying throughout by n

$$(n\,V_g + n\,V_a - V_g - V_a)\,b_1 = n\,V_g - V_g.$$
$$\therefore \quad (V_g + V_a)\,(n-1)\,b_1 = V_g\,(n-1)$$

and $$b_1 = \frac{V_g}{V_g + V_a} = \frac{\text{genetic variance}}{\text{total variance within plots}}.$$

Substituting for b_1 in equation (III)

$$\left\{\frac{1}{n}(V_g + V_a) + V_b\right\} b_2 = V_g - V_g \frac{(V_g + V_a + V_b)}{V_g + V_a}$$

$$= -\frac{V_g V_b}{V_g + V_a},$$

$$\therefore \quad b_2 = -\frac{V_g V_b}{(V_g + V_a)\left\{\frac{1}{n}(V_g + V_a) + V_b\right\}},$$

and $$b_1 + b_2 = \frac{V_g}{V_g + V_a} - \frac{V_g V_b}{(V_g + V_a)\left\{\frac{1}{n}(V_g + V_a) + V_b\right\}}$$

$$= \frac{V_g (V_g + V_a + n V_b) - n V_g V_b}{(V_g + V_a)(V_g + V_a + n V_b)}$$

$$= \frac{V_g (V_g + V_a)}{(V_g + V_a)(V_g + V_a + n V_b)} = \frac{V_g}{(V_g + V_a + n V_b)}$$

$$= \frac{\text{genetic variance}}{\text{total variance between plots}}.$$