

THE PHYSIOLOGICAL CONSEQUENCES OF POLYPLOIDY¹

II. THE EFFECT OF POLYPLOIDY ON VARIABILITY IN THE TOMATO

BY A. C. FABERGÉ

John Innes Horticultural Institution

INTRODUCTION

THE data of the experiments described in an earlier paper (Fabergé, 1936) in which the effect of polyploidy on growth and total size was considered, lend themselves very well to an analysis of variability. One of these experiments, called the 1935 growth-rate experiment, was in fact mainly designed for the purpose. Full experimental details are given in that paper, and will not be repeated here.

Lindstrom (1935) studied the frequency distribution of fruit weight in diploid and tetraploid F_2 populations of *Solanum racemigerum* × Tomato. According to Lindstrom, the tetraploid condition results in:

- (a) a reduction of fruit weight;
- (b) a reduction in relative variability;
- (c) the disappearance of positive skewness on an arithmetic scale.

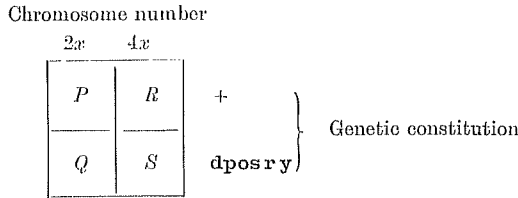
When the data are transposed into a logarithmic metric, the positive skewness of the diploid vanishes, but the tetraploid becomes negatively skew. The reduction in relative variability is attributed to tetrasomic segregation ratios, and a general dominance of quantitative factors—in this case for small size. In this way a lesser proportion of recessive phenotypes appear in a tetraploid F_2 as compared with a diploid, and a narrower frequency distribution results. Lindstrom interprets his skewness result by assuming that gene action is in this case mainly additive, and not geometric, in the sense of Rasmusson (1933). Fisher's (1935, 1936) methods of analysis of variance have been used in analysing the present data, and his terminology is used throughout.

¹ In part adapted from a thesis accepted for the degree of Ph.D. of the University of London.

EXPERIMENTAL DATA

1934 *experiment*

In the 1934 experiment four inbred populations (six generations of selfing) *P*, *Q*, *R* and *S* were used. These may be classified as follows:



The two tetraploids had been obtained by somatic doubling in 1932, and the corresponding diploids came from selfing the same individual in that year. The factors are called by the symbols of MacArthur (1931); a full description of them is given in his paper. The plants were grown in water culture in a greenhouse, and nine samples were taken at four-day intervals, covering a month of growth. Twelve plants were taken from each line at each time. Eight randomized blocks were used, but, as was seen in the previous paper (Fabergé, 1936), the sum of squares for blocks is not significant and need not be taken into account. Two plants were accidentally lost in sample 3 of line *P*, so that there are in all 429 degrees of freedom instead of 431. The analysis of variance is then as follows:

				Degrees of freedom
Lines...	3	
Times	8	
LT	24	
Individuals	...			
LI		} 394 = Error
TI		
LTI		
Total...	429	

Four measurements were taken on each individual plant; dry and fresh weight of both roots and shoots. They are very highly correlated, and give identical results; only the dry weight of shoots need be considered here. Logarithms to the base 10 of the weight are used in the computations, 0.1 mg. being a unit. A consequence of using a logarithmic metric is that the mean square becomes independent of the mean, and this supplies a *measure* of relative variability, in the same sense as Pearson's coefficient of variability.

It is with the analysis of the 394 degrees of freedom for error that we are concerned here. This is done by means of a 4 × 9 table (Table I).

Each item in this table is the sum of the squares of deviations from a sample of twelve. Each sample is determined from its position in the table with respect to the four lines and nine times. The marginal totals supply the required information.

The four line totals show that the two diploids *P* and *Q* are much more variable than the two tetraploids *R* and *S*. This difference is highly significant:

TABLE I*
Decomposition of error of 1934 experiment

Times	Lines				Total
	<i>P</i>	<i>Q</i>	<i>R</i>	<i>S</i>	
1	0.2756	0.3200	0.0655	0.1941	0.8551
2	0.3663	0.4148	0.2796	0.2016	1.2623
3	0.1440	0.1932	0.1569	0.1363	0.6307
4	0.1284	0.3078	0.2168	0.1213	0.7743
5	0.4485	0.5189	0.1726	0.1415	1.2815
6	0.6895	0.4916	0.0728	0.1297	1.3835
7	0.2860	0.3670	0.1197	0.3154	1.0880
8	0.3810	0.4397	0.1526	0.3179	1.2912
9	0.2855	0.2787	0.1322	0.1556	0.8521
Total	3.0048	3.3318	1.3688	1.7134	9.4188

* Numbers rounded off from computations made with more decimals. Hence some totals do not check in last figure.

TABLE II

Source of variability	D.F.	Sum of squares	Mean square	<i>z</i>	0.01 point in the distribution of <i>z</i>
<i>P</i> + <i>Q</i>	196	6.3366	0.0323	0.3654	0.1662
<i>R</i> + <i>S</i>	198	3.0821	0.0156		

The other classification, with respect to genetic constitution, only gives a very small difference which is not at all significant:

TABLE III

Source of variability	D.F.	Sum of squares	Mean square	<i>z</i>	0.05 point in the distribution of <i>z</i>
<i>P</i> + <i>R</i>	196	4.3736	0.0223	0.0664	0.1175
<i>Q</i> + <i>S</i>	198	5.0452	0.0255		

Turning now to the totals on the right-hand margin of Table I, it is clear that there is no general increase of error with time, on a logarithmic scale. We may now test whether the variations in this series of errors can be attributed to random sampling. The general test for the heterogeneity of a series of variances is given in an Appendix to this paper by Mr W. L. Stevens (p. 398).

The variance of the series of errors is:

$$\frac{S(e - \bar{e})}{8} = 0.0748.$$

This value is compared with the variance of a theoretical χ^2 distribution :

$$88 \left(\frac{9.4187}{394} \right)^2 = 0.0496$$

(a slight correction has been made because entry 3*P* is based not on 11, but on 9 degrees of freedom). The ratio of these two variances multiplied by 8 gives a χ^2 value of 12.064, based on 8 degrees of freedom. According to Fisher's table of the distribution of χ^2 , the probability of exceeding it by random sampling lies between 0.2 and 0.1. One may thus consider the variations of this column as being due to random sampling. Since logarithms were used, this result shows that the variability of the populations may be considered as being only the W_o component of Blackman's growth equation :

$$W_t = W_o e^{rt}$$

where W_t = weight at time t , W_o = initial weight and r = relative rate or efficiency index (Blackman, 1919).

Embryo weights of lines P, Q, R, and S

In the first sample of the 1934 experiment, the plants had about five times the weight of the embryos. It appeared important to see if the reduction of variability associated with the tetraploid condition could be found in the embryos themselves.

Fifty embryos of each of the lines *P*, *Q*, *R* and *S* were dissected out, dried and weighed individually. The methods are fully described in the previous paper (Fabergé, 1936). In addition to variability, it was desired to test the normalities of the distributions, in view of Lindstrom's results. Lindstrom, as was seen before, found that the positive skewness of the distribution of diploid fruit weights disappears in the tetraploid. The cubes and fourth powers, together with the sums and the squares of the observations were accumulated by the method of Progressive Digitizing on Hollerith punched card machines (Baehne, 1935). I am very much indebted to the British Tabulating Machine Company for letting me use their apparatus.

An analysis of the absolute variability, on an arithmetic scale, is given in Table IV.

The difference in *absolute* variability between the diploid and tetraploid embryos is not significant, though it is in the same direction as in the growing population. It is also clear that the greater part of the difference is due to the very high variability of the line *Q*.

It was seen in the previous paper (Fabergé, 1936) that tetraploid

embryos are about 30 per cent heavier than diploid ones, it is consequently necessary to use a measure of *relative* variability. I am much indebted to Mr W. L. Stevens for pointing out a method for getting an estimate of the variance of $\log x$ without having to recalculate the sums

TABLE IV

Analysis of absolute variability: embryo weights

	D.F.	S. of S.	M.S.	D.F.	M.S.	z	0.05 z
Total	196	94267	480.96				
<i>P</i> 2 <i>x</i>	49	21888	446.69	98	540.82	0.1244	0.1662
<i>Q</i> 2 <i>x</i>	49	31112	634.94				
<i>R</i> 4 <i>x</i>	49	20459	417.53	98	421.09		
<i>S</i> 4 <i>x</i>	49	20808	424.65				

of squares after transforming into logarithms. It represented a saving of time, since the necessary third and fourth powers were already available for the tests of normality. The numerical values of the terms of this transformation are given in Table V, where S_2 , S_3 , and S_4 represent the sums of the 2nd, 3rd and 4th powers of deviations from the mean respectively. The series of course continues, but only the first three terms are used here.

TABLE V

$$\frac{1}{\bar{x}^2} \cdot \frac{1}{n-1} \cdot S_2 - \frac{1}{\bar{x}^3} \cdot \frac{1}{n-1} \cdot S_3 + \frac{11}{12} \cdot \frac{1}{\bar{x}_4} \cdot \frac{1}{n-1} \cdot S_4 - \frac{1}{4} \cdot \frac{1}{\bar{x}^4} \cdot \frac{n-1}{n^2} \left(\frac{1}{n-1} S_2 \right)^2$$

<i>P</i>	0.047045	+0.005242	+0.004655
<i>Q</i>	0.063851	-0.014742	+0.011489
<i>R</i>	0.021886	+0.000244	+0.001032
<i>S</i>	0.027309	+0.002036	+0.001782

It will be noted that by taking the square root of the first term and multiplying by 100, Pearson's coefficient of variability is obtained.

The result of the transformation is given in Table VI. Under "growing population" are given the variances from Table I multiplied by $(\text{Log}_e 10)^2 = 5.30197$, since these were calculated on logarithms to the base 10.

TABLE VI

Lines	Embryos		Growing population		$V(\rho)/V(e)$
	D.F.	Logarithmic variance	D.F.	Logarithmic variance	
<i>P</i> 2 <i>x</i>	49	0.05694	97	0.16436	2.886
<i>Q</i> 2 <i>x</i>	49	0.06060	99	0.17868	2.949
<i>R</i> 4 <i>x</i>	49	0.02316	99	0.07317	3.159
<i>S</i> 4 <i>x</i>	49	0.03113	99	0.09172	2.947

In the last column is given the ratio of relative variability of the population to that of the embryos. From this it is clear that the varia-

bilities of the growing populations are very closely proportional to those of the corresponding embryos, though they are about three times as great. Thus the variability which, as was seen before, does not increase during growth, does so during germination.

The difference between the relative variability of diploid and tetraploid embryos, unlike the difference in absolute variability, is highly significant, as the following table shows:

TABLE VII

Lines		D.F.	Relative variability	z	0.01 pt. of z
$P+Q$	$2x$	98	0.05877	0.3861	0.2326
$R+S$	$4x$	98	0.02715		

The tests of normality on the embryo populations are given in Table VIII. g_1 and g_2 are used (Fisher, 1936); the standard errors appended to them are large sample standard errors.

TABLE VIII

Tests of normality: embryo weights

Line	g_1	S.E. of g_1	g_2	S.E. of g_2
P	-0.4932		-0.5936	
Q	+0.8772	± 0.3366	+0.2880	± 0.4381
R	-0.0725		-0.5314	
S	-0.4331		-0.2403	

There is clearly no consistent difference in skewness between diploids and tetraploids such as Lindstrom found in the case of fruit weights. It should be noted on the other hand that tests of skewness with fifty individuals are not very sensitive. Line Q alone departs significantly from normality. It is interesting to note that the outstandingly high absolute variability of this line vanished in a logarithmic metric, and that this is due to the correction introduced by the third power term in Table V. Had the coefficient of variability been used instead of the logarithmic variance, the measure of relative variability would have been affected by the skewness.

None of the four populations depart significantly from normality symmetrically.

1935 *experiment*

In the 1935 experiment ten populations were used, which may be classified as follows:

$$\text{Diploid } \dots \dots \left\{ \begin{array}{l} \text{Pure lines } \dots \dots \left\{ \begin{array}{l} A \\ I \\ J \end{array} \right. \\ E_2 \dots \dots \dots \left\{ \begin{array}{l} F \\ G \end{array} \right. \end{array} \right.$$

$$\text{Tetraploid} \dots \left\{ \begin{array}{l} \text{Pure lines} \dots \dots \dots \left\{ \begin{array}{l} B \\ C \\ D \end{array} \right. \\ F_2 \dots \dots \dots \left\{ \begin{array}{l} E \\ H \end{array} \right. \end{array} \right.$$

Full descriptions of these lines are given in the previous paper. It need only be noted here that in the case of *G* and *H*, the same individual F_1 plant is the parent of both, and that *A* and *B* came from selfing the same individual in 1932.

The experiment is arranged in four randomized blocks. Eight samples at intervals of one week were taken, 12 plants from each line (three from each block). The first three samples came from a greenhouse, the remaining five from outside. Only the dry weight of shoots was recorded.

The theoretical structure of the experiment may then be represented as follows:

TABLE IX

Criteria of classification and their interactions	Degrees of freedom	Items appearing as such in the main analysis	Items making up absolute error	Items making up residual error
Lines	9	+	.	.
Times	7	+	.	.
Individuals	2	.	+	.
Blocks	3	.	.	+
LT	63	+	.	.
LI	18	.	+	.
LB	27	.	.	+
TI	14	.	+	.
TB	21	.	.	+
IB	6	.	+	.
LTI	126	.	+	.
LTB	189	.	.	+
LIB	54	.	+	.
TIB	42	.	+	.
LTIB	378	.	+	.
Total	959	.	640	240

It is the absolute error which supplies the information on variability. It is based on 640 degrees of freedom, and consists, as will be seen from the table, of *I* and of all the interactions into which *I* enters.

The analysis is carried out in essentially the same way as in the case of the 1934 experiment. A slight complication is introduced by the use of blocks which, unlike those of the 1934 experiment, are highly significant. Of the 11 degrees of freedom in each sample, three are due to blocks. Each entry of the decomposition of error table is the sum of squares based on the remaining 8 degrees of freedom. The decomposition of error is given in Table X.

This resembles the analysis obtained in 1934 very closely. It is clear from the right-hand column of totals that there is no general increase of

error with time on a logarithmic scale. But the variation in this column cannot be accounted for by assuming that it is the result of random sampling, as was the case in the 1934 experiment. The test, when carried out in the same way, gives a χ^2 value of 43.1561 for only 7 degrees of free-

TABLE X
Decomposition of error of 1935 experiment

Times	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	Total
1	0.53535	0.34089	0.43945	0.21799	0.54641	0.13425	0.23003	0.22113	0.03303	0.33970	3.03823
2	0.34959	0.32653	0.57519	0.09569	0.52286	0.20765	0.33430	0.11275	0.15083	0.61205	3.28744
3	1.06650	0.15995	0.14537	0.43001	0.43897	0.62372	1.33180	0.12104	0.05725	1.60500	6.28021
4	0.37193	0.08264	0.13118	0.04075	0.16493	0.35316	0.54804	0.10172	0.09952	0.06498	1.95885
5	0.10138	0.13716	0.07881	0.17507	0.27852	0.10927	1.04003	0.56744	0.11544	0.30671	3.90983
6	0.18104	0.11857	0.15603	0.27812	0.17632	0.11342	0.92087	0.33547	0.36146	0.06965	3.01395
7	0.78500	0.22968	0.21605	0.31651	0.16680	0.85123	0.13013	0.81581	0.27220	1.17267	4.95698
8	0.15377	0.26167	0.04905	0.30754	0.42499	0.51180	0.14836	0.61058	0.19737	0.17847	2.84360
Total	3.54756	1.95709	2.09203	1.86168	2.71980	2.90450	4.68356	2.88594	1.28710	4.34983	28.28909

dom, which is very significantly above expectation. No doubt this is due to uncontrolled factors inevitable in an outdoor experiment, which also covered a much longer period of growth, viz. 2 months. There is no reason to suspect the sampling technique adopted.

The comparison of diploid and tetraploid populations as regards variability is as follows:

TABLE XI

Variance within	D.F.	s. of s.	M.S.	<i>z</i>	0.01 <i>z</i>
Diploids	320	16.77255	0.052414	0.1880	0.1303
Tetraploids	320	11.51654	0.035989		

The difference is highly significant, and is in entire agreement with the results of the 1934 experiment.

It is of interest to compare separately the F_2 populations *G* and *H*, and also the pure lines *A* and *B*. The former pair comes from selfing the same F_1 individual, the latter from selfing the same individual in 1932. These comparisons are given below:

TABLE XII

Variance within	D.F.	s. of s.	M.S.	<i>z</i>	0.01 <i>z</i>
F_2 <i>G</i>	64	4.6836	0.07318	0.2421	0.2931
F_2 <i>H</i>	64	2.8859	0.04509		0.05 <i>z</i> 0.2072

TABLE XIII

Variance within	D.F.	s. of s.	M.S.	<i>z</i>	0.01 <i>z</i>
Pure line <i>A</i>	64	3.5476	0.05543	0.2974	0.2931
Pure line <i>B</i>	64	1.9571	0.03058		

In the former case, the difference is significant, although it does not reach the 0.01 level of probability. In the latter case the difference is highly significant.

The evidence from these two pairs of lines is more critical than from the remainder of the experiment, since they are very strictly comparable. It is clearly apparent that genetic constitution is capable of strongly influencing the amount of variability of a line.

It will be seen from the main decomposition of error table that the least variable of all lines is a diploid, the pure line *I*. This difference is highly significant, as the following test shows:

TABLE XIV

Variance within	D.F.	s. of s.	M.S.	<i>z</i>	0.01 <i>z</i>
Pure line <i>I</i>	64	1.2871	0.02011		
Pure lines <i>A-H, J</i>	576	27.0020	0.04688	0.3675	0.2005

No reason can be suggested for this apparently exceptional behaviour. The line *I* is a genetic stock called "Chinaman". It is known to have been cultivated since about 1860 in Australia, and is homozygous for several recessives.

The decomposition of error can also yield information on the comparative variability of pure lines and F_2 populations:

TABLE XV

Variance within	D.F.	s. of s.	M.S.	<i>z</i>	0.01 <i>z</i>
Pure lines	384	15.0953	0.03931		
F_2	256	13.1938	0.05154	0.1354	0.1313

Thus F_2 's are more variable than pure lines. To show that this is not due to any general increase of variability of F_2 's with time as compared with pure lines, the mean squares at each time are given below.

TABLE XVI

Times	Mean square	
	P.L.	F_2
1	0.0397	0.0354
2	0.0440	0.0368
3	0.0784	0.0786
4	0.0165	0.0365
5	0.0191	0.0624
6	0.0306	0.0483
7	0.0624	0.0614
8	0.0239	0.0530

Fruit weights

In the autumn of 1935 an investigation was undertaken which is similar in character to the one described by Lindstrom (1935) and discussed in an earlier part of this paper.

The experiment was not planned sufficiently long in advance for a proper field technique to be adopted; use was made of Dr Sansome's genetic populations which were intended for qualitative work only. It is confidently believed, however, that the main results are not in any way invalidated by this defect of experimental design.

Two F_2 populations were used—a diploid and a tetraploid. Both were obtained by selfing the same F_1 individual, the tetraploid being obtained by somatic doubling of part of that same individual. The parents were the stock "Chinaman" (line *I* of the previous experiment) and a pure line containing a new recessive factor making all parts of the plant very hirsute.

Sixty-eight plants of each of the two populations, growing in two long parallel rows, were used. The land was, so far as could be judged from the remaining crop, very uniform. Ten fruits were weighed from each of the 136 plants. In order to avoid bias, they were taken starting from the bottom of the plant and working outwards on each truss. Only fruits obviously injured, of which there were very few, were omitted in this order of picking. Fruit weights were recorded to the nearest gram.

An analysis of the results is given in Table XVII.

TABLE XVII
Analysis of variance of fruit weight

Variance	D.F.	2x		4x	
		s. of s.	M.S.	s. of s.	M.S.
Total	679	1162717.28	1712.40	323207.29	476.00
Between plants	67	325903.98	4864.24	138110.09	2061.34
Within plants	612	836813.30	1367.34	185097.20	302.45

Actual numbers, not logarithms, are used here.

It is clear that there is the same reduction of variability as was found by Lindstrom, and it is of the same order of magnitude. The total amount of variability is rather greater in our case than in his—a remarkable circumstance, since the parents of Lindstrom's F_2 differed much more in respect of fruit size than the ones used here.

In terms of the conventional coefficient of variability, the two sets of data are as follows:

	Lindstrom's data	Present data
2x	36.5	50.23
4x	29.4	40.53

An analysis of relative variability is given in Table XIX. By variability is designated the statistic

$$\frac{1}{x^2} \cdot \frac{1}{n-1} S(x-\bar{x})^2.$$

This corresponds to the first term of the transformation given in Table V, and its square root ($\times 100$) is the coefficient of variability.

TABLE XIX

Relative variability

Source of variability	D.F.	2x	4x	z	Points of z	
					0.01	0.0893
Total	679	0.2520	0.1704	0.1956	0.01	0.0893
Between plants	67	0.7158	0.7108	0.0035	0.05	0.2025
Within plants	612	0.2012	0.1043	0.3285	0.01	0.0941

The most striking thing in this table is that the entire reduction in variability is in the "within plants" component. The very small difference between plants is not at all significant, as the z tests show. This result has very important consequences on the interpretation of other data presented here.

These fruit measurements also show that polyploidy results in a reduction of fruit size, as has been found before by Jørgensen (1928) and Lindstrom (1935).

TABLE XX

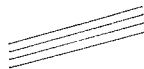
Fruit weight

	Mean weight in gm.	Difference	Standard error
2x	82.44		
4x	53.85	28.59	2.83

It should be noted that although the mean fruit weight is about 16 \times greater in our case than in Lindstrom's, the proportional reduction resulting from polyploidy is about the same.

DISCUSSION

It appears from the data presented that polyploidy consistently results in a diminished variability. The results of the growth-rate experiments clearly show that segregation of factors controlling growth rate does not occur to an appreciable extent. The populations may then be represented on a logarithmic metric as consisting of parallel lines:



and not of lines at an angle to one another:



In other words, the whole of the variation is in the W_0 component of the exponential equation

$$W_t = W_0 e^{rt}$$

which Blackman (1919) has applied to the growth of plants.

Significant differences between the variabilities of diploids and tetraploids are found on each occasion; it follows that this is a difference in the variabilities of initial weights, and not of growth rates. The embryo data fully confirm this, since it was shown that the lesser variability of tetraploids is already present there. The variabilities of the embryos were very closely proportional to those of the growing populations, though the latter were about three times greater. It is difficult to say, in the absence of exact knowledge about growth during germination, what the mechanism of this accentuation of variability is. It is reasonable to suppose that the greater part is due to differences in the time when germination begins. In the previous paper it was shown that tetraploids (having 30 per cent heavier embryos) germinate more slowly, but the data given here show that *within lines*, larger embryos in some way gain proportionately more.

If Lindstrom's explanation of the reduced variability of tetraploids is to be applied to this case, it would be necessary to assume that all genes determining total plant size have already finished acting at the time the seed is ripe. This seems very unlikely. The absence of any consistent change in skewness is also against Lindstrom's view.

The evidence from fruit weights is quite categorical. Since the difference in variability is only between *different parts of the same individual*, it cannot be due to genetic differences between individuals. This at once makes it impossible to apply Lindstrom's explanation, which requires tetrasomic segregation ratios and a dominance of quantitative genes. Lindstrom, as was already seen, had deduced his hypothesis from fruit weight data precisely similar to those given here, and which in fact very closely resemble his own quantitatively. It is particularly regrettable that the methods of Analysis of Variance were not used in Lindstrom's case, for it is only by separating the total sum of squares into two components, ascribable respectively to "within" and "between" plants, that the true nature of the variation can be demonstrated.

The almost complete absence of difference of variability "between" plants must mean that a bigger plant does not produce heavier individual fruits. It may of course produce more fruits. It was seen that F_2 populations were rather more variable than pure lines in the 1935 experiment. Much weight cannot be attached to this, since the populations were not in any sense comparable.

A priori there is no obvious reason why polyploidy should reduce the variability at all, and it is therefore reasonable to attribute both the diminished variability of whole plants and of fruits to the same cause. The analysis of the results on fruits clearly shows that the component of variability which is reduced by chromosome doubling is not due to genetic differences between the objects measured. We must therefore be dealing with non-heritable variation. If this applies equally to whole plants, the difficulty of assuming that all quantitative factors have already finished acting when the seed is ripe is removed, and the absence of any change in skewness accounted for. It has been customary, more particularly since Johanssen's classical experiments with beans (Johanssen, 1909) to divide all variation into two classes: genetic and environmental. Now, the kind of variation dealt with here, the extent of which is diminished by tetraploidy, is clearly not genetic. But to say that it is environmental requires a discussion of what is meant by environment in this case.

The decompositions of error show that, in the case of whole plants at any rate, one must not think of an environment acting throughout development to produce differences. There is no information in the data presented here as to when variability arises in fruits, but it is of interest to note that Houghtalling (1935) has shown that the size of tomato fruits is already fully determined at a very early stage (open flowers). Thus here again it would appear that there is no question of an environment acting continuously throughout development, but only during the formation of early meristems.

Genes determine the form and size of the organism. During their action, circumstances may arise under which two genes fail to have full effect, but four are sufficient for a complete or more complete action. It is suggested that a process of this kind will best account for the effect observed. In any case this must take place in the very earliest stages of development. The existence of such a mechanism of "strengthening" of gene action by the increase of the number of genes cannot of course be regarded as proved by the data presented here. It is merely put forward as plausible. The subject is too new, and the technique too little developed for a definite answer to be given without much more extensive investigations.

It remains to point out that the idea is not unrelated to Timoféeff-Ressovsky's treatment of the intensity of gene manifestation (Timoféeff-Ressovsky, 1934). The "intensity" of different allelomorphs is classified into "Penetranz" or the probability of the gene producing a visible effect at all, and "Expressivität", which is the degree of its manifestation. Both are in any case highly correlated. In this terminology, one might say that either or both the Penetranz and the Expressivität of quantitative genes is increased by an increase in their absolute number.

SUMMARY

1. Experiments described in a previous paper, comparing diploid and tetraploid tomatoes with regard to size and growth, are analysed from the point of view of variability.

2. New data on fruit weight in diploid and tetraploid tomatoes are given.

3. Tetraploids are found to be consistently less variable than diploids. When the weight of whole plants is considered, it is shown that this reduction of relative variability lies entirely in the W_0 term of Blackman's growth equation $W_t = W_0 e^{rt}$. In fruits, the reduction of variability is between fruits on the same plant, and not between different plants.

4. It is considered that this diminished variability cannot be accounted for by a direct genetic effect involving segregation, as had been suggested by other workers.

5. An alternative hypothesis is proposed. It is assumed that doubling the number of genes results in an increased probability of the action of quantitative factors. This is reflected in the greater physiological stability of early meristems.

6. The fact that tetraploidy results in diminished fruit size is confirmed.

ACKNOWLEDGEMENTS

I am much indebted to Dr F. W. Sansome for supplying the material used in this work, and for advice in the preparation of the paper. To Prof. R. A. Fisher for much valuable advice, and for his kindness in allowing me to do the statistical work in the Galton Laboratory. To Prof. J. B. S. Haldane for reading the manuscript and helpful criticism. To Dr E. Ashby for allowing me to cite some of his unpublished data.

REFERENCES

- ASEBY, E. (1930). "Studies in the inheritance of physiological characters. I. A physiological investigation of the nature of hybrid vigour in maize." *Ann. Bot.* **44**, 457-67.
- (1932). "Studies in the inheritance of physiological characters. II. Further experiments upon the basis of hybrid vigour, and upon the inheritance of efficiency index and respiration rate in maize." *Ibid.* **46**, 1007-33.
- BAEENE, G. W. (edited by) (1935). *Practical Applications of the Punched Card Method in Colleges and Universities*. New York (Columbia Univ. Press).
- BLACKMAN, V. H. (1919). "The compound interest law and plant growth." *Ann. Bot.* **33**, 353-60.
- FABERGÉ, A. C. (1936). "The physiological consequences of polyploidy. I. Growth and size in the tomato." *J. Genet.*
- FISHER, R. A. (1935). *The Design of Experiments*. Pp. 252. Edinburgh (Oliver and Boyd).
- (1936). *Statistical Methods for Research Workers*. 6th ed. Pp. 336. Edinburgh (Oliver and Boyd).
- HOUGHTALLING, H. B. (1935). "A developmental analysis of size and shape in tomato fruits." *Bull. Torrey Bot. Club* **62**, 243-51.
- JØRGENSEN, C. A. (1928). "The experimental formation of heteroploid plants in the genus *Solanum*." *J. Genet.* **19**, 133-211.
- JOHANNSEN, W. J. (1909). *Elemente der exakten Erblchkeitslehre*. Jena.
- LINDSTROM, E. W. (1935). "Segregation of quantitative genes in tetraploid tomato hybrids as evidence for dominance relations of size characters." *Genetics*, **20**, 1-11.
- MACARTHUR, J. W. (1931). "Linkage studies with the tomato." *Trans. Roy. Canad. Inst.* **18**, 1-19.
- RASMUSSEN, J. (1933). "A contribution to the theory of quantitative inheritance." *Hereditas*, **18**, 245-61.
- TIMOFÉEFF-RESSOVSKY, N. W. (1934). "Über den Einfluss des genotypischen Milieus und der Aussenbedingungen auf die Realisation des Genotyps. Genmutation vti bei *Drosophila funebris*." *Nachr. Ges. Wiss. Göttingen Phys-Math. Kl.* **6**, 53-106.

APPENDIX

HETEROGENEITY OF A SET OF VARIANCES

By W. L. STEVENS

Galton Laboratory, University College, London

We shall first consider a general problem, of which the test of heterogeneity of a set of estimated variances is a special case.

Let $x_r, r=1, 2 \dots k$, be k quantities normally distributed with known variances σ_r^2 about unknown means μ_r , i.e.,

$$x_r = \mu_r + \sigma_r \eta_r$$

where η_r is normally distributed about zero with unit standard deviation. Then it is required to test the hypothesis that the means μ_r are equal for all r .

If the hypothesis is true, an efficient estimate of the common mean would be

$$\begin{aligned} m &= \frac{\sum_{r=1}^k \left(\frac{x_r}{\sigma_r^2} \right)}{\sum_{r=1}^k \left(\frac{1}{\sigma_r^2} \right)} \\ &= \mu + \frac{\sum \left(\frac{\eta_r}{\sigma_r} \right)}{\sum \left(\frac{1}{\sigma_r^2} \right)}. \end{aligned}$$

Hence

$$\begin{aligned} \sum \left\{ \frac{1}{\sigma_r^2} (x_r - m)^2 \right\} &= \sum \left\{ \eta_r - \frac{1}{\sigma_r} \cdot \frac{\sum \left(\frac{\eta_r}{\sigma_r} \right)}{\sum \left(\frac{1}{\sigma_r^2} \right)} \right\}^2 \\ &= \sum_{r=1}^k \eta_r^2 - \frac{\left\{ \sum_{r=1}^k \left(\frac{\eta_r}{\sigma_r} \right) \right\}^2}{\sum_{r=1}^k \left(\frac{1}{\sigma_r^2} \right)}. \end{aligned}$$

Since the quantity $\sum \left(\frac{\eta_r}{\sigma_r} \right) / \sqrt{\sum \left(\frac{1}{\sigma_r^2} \right)}$ is a linear function of the η 's, with the sum of the squares of the coefficients equal to unity, it follows that

$$\sum \left\{ \frac{1}{\sigma_r^2} (x_r - m)^2 \right\} = \chi^2_{[k-1]}$$

i.e., is distributed in a χ^2 of $k-1$ degrees of freedom(1). The table of χ^2 may therefore be used to make a test of significance.

Application to a set of estimated variances. The hypothesis to be tested is that k sets of quantities x_{rt} ($r=1, 2 \dots k, t=1, 2 \dots n_r$) are distributed with a common variance σ^2 , respectively about the means $\mu_1, \mu_2 \dots \mu_k$, the values of which are unknown and need not be estimated.

For each set we have the estimate of variance

$$s_r^2 = \frac{\sum_{t=1}^{n_r} (x_{rt} - \bar{x}_r)^2}{n_r - 1}$$

which has a mean σ^2 , and variance $2\sigma^4/(n_r-1)$.

Following the previous argument, we find that the best estimate of σ^2 is

$$\begin{aligned} & \frac{\sum_{r=1}^k \left\{ \frac{s_r^2 (n_r - 1)}{2\sigma^4} \right\}}{\sum_{r=1}^k \left(\frac{n_r - 1}{2\sigma^4} \right)} \\ &= \frac{\sum \{ (n_r - 1) s_r^2 \}}{\sum (n_r - 1)} = s^2 \end{aligned}$$

as would be found by the usual method. Hence

$$\sum \left\{ \frac{n_r - 1}{2\sigma^4} (s_r^2 - s^2) \right\}$$

is distributed very nearly in a χ^2 of $k-1$ degrees of freedom. As however σ^2 is unknown, it is necessary to approximate by putting $\sigma^2 = S^2$, and we find that

$$\frac{1}{2S^4} \sum_{r=1}^k \left\{ (n_r - 1) (s_r^2 - S^2)^2 \right\}$$

is approximately in a χ^2 of $k-1$ degrees of freedom. Since this formula is sensitive to differences between the variances it is suitable for testing the hypothesis that the variances are equal.

It may be noted that in large samples the above test tends to equivalence to the one given by Pearson and Neyman (1931) based on the ratio of the weighted geometric and arithmetic means(2).

REFERENCES

- (1) FISHER, R. A. (1936). *Statistical Methods for Research Workers*. 6th ed. Table of χ^2 , p. 118.
- (2) NEYMAN, J. & PEARSON, E. S. (1931). "On the problem of k samples." *Bulletin de l'Académie Polonaise des Sciences et des Lettres* (Série A, Sciences Mathématiques), p. 460.