ENVIRONMENTAL AND GENOTYPE-ENVIRONMENTAL COMPONENTS OF VARIABILITY

I. INBRED LINES

L. BUCIO ALANIS*

Department of Genetics, The University, Birmingham

1. INTRODUCTION

CONTINUOUS variation is of major importance in evolution as well as in the more practical aspect of breeding for economic traits. A full description of phenotypic variation requires three main components *viz.*, genetic, environmental and genotype-environmental interaction, that is:

$$P = G + E + GE.$$

Much work has been directed towards an understanding and description of genetic variability. Less effort has been devoted to the environmental component and still less to the interactive effect of the genotype and the environment. The present paper is primarily concerned with these two latter sources of variation.

The data used is taken from the results of an experiment on *Nicotiana rustica* initiated in 1946 by Professor Mather and thereafter continued for some years by himself and his colleagues (table 1). The experiment included two inbred lines P_1 and P_5. (Mather and Vines, 1952) and some of the various generations which can be raised from crosses between them, such as the F_1, F_2, backcrosses, etc. In all years flowering time from an arbitrary date and final height in inches were recorded. But for the purposes of the present investigation, only the data referring to the final height of the two inbred lines, P_1 and P_5, has been considered. These two lines have been propagated from year to year by the artificial selfing of individual plants. The experiment was grown at the John Innes Institute in London from 1946 to 1948, and from 1950 to 1964 at the University of Birmingham. From 1946 to 1958 the seeds were planted in boxes in the glasshouse and after 25 to 30 days the plants were transplanted to the field. From 1959 onwards the seeds were sown in individual pots made of compressed organic matter. After 25 to 30 days the intact pots were transferred to the field; thus the plants did not suffer any effect of transplantation and in general developed much better.

From 1946 to 1958 the experiments consisted of two blocks, within each block the inbred lines and other generations derived from them being placed in plots containing five plants each. From 1959 to 1964

the experiments consisted of two blocks, the individual plants of the various generations being placed completely at random within each block.

2. RESULTS AND DISCUSSION

The mean final heights of the two inbred varieties of *N. rustica*, P_1 and P_5 when grown under different environments as represented by seasons and locations are given in table 1 and fig. 1. From these we can see that two inbred lines show different responses to the changing

TABLE 1

Plant height in inches of the two inbred lines P_1 *and* P_5 *of* N. rustica
1946-1964 experiments

Investigator	Year	Location	Number of obs. n_1	Mean height P_1	Number of obs. n_5	Mean height P_5
Mather	1946	London	50	39·36	50	49·10
Mather	1947	London	60	39·60	50	50·14
Mather	1948	London	50	39·48	49	48·16
Mather	1950	Birmingham	47	37·13	47	41·17
Mather	1951	Birmingham	49	40·20	50	42·94
Breese	1952	Birmingham	50	38·34	49	37·37
Mather	1953	Birmingham	50	38·14	50	40·54
Opsahl	1954	Birmingham	50	41·90	50	44·68
Breese	1956	Birmingham	20	42·15	20	42·50
Arlett	1957	Birmingham	50	35·84	50	32·86
Arlett	1958	Birmingham	40	44·45	40	47·43
Arlett	1959	Birmingham	20	36·75	13	44·62
Hill	1960	Birmingham	40	48·70	40	60·73
Hill	1961	Birmingham	36	47·31	36	57·69
Hill	1962	Birmingham	39	45·23	39	53·49
Bucio	1964	Birmingham	28	46·36	30	60·03

environment, although an interpretation of the nature of the different responses is not obvious. Nevertheless, some useful information can be obtained, for example:

(*a*) P_5 is, on average, taller than P_1 although in certain specific environmental conditions the reverse is true.

(*b*) P_1 shows more stability of performance in plant height over environments than P_5.

An analysis of variance could be used to estimate the relative magnitudes of the average values of σ_e^2 and σ_g^2, the environmental and genotype-environmental variance components respectively, but here such estimates will be obtained in a slightly different manner in order to show the role which each component plays in the phenotypic expression. The estimation of such variance components is important because it gives information regarding the possibilities of selection and such an analysis is a convenient starting point in a breeding programme. Nevertheless we need in addition more knowledge of the nature of the

first degree statistics of the environmental and genotype-environmental parameters.

We could start by defining the best genotype as that which has

(*a*) Highest performance over environments.

(*b*) Highest stability of performance (lowest variance over the possible environments).

This definition is a desirable one, but as will be seen later, it by no means implies that highest performance and stability go hand in hand.
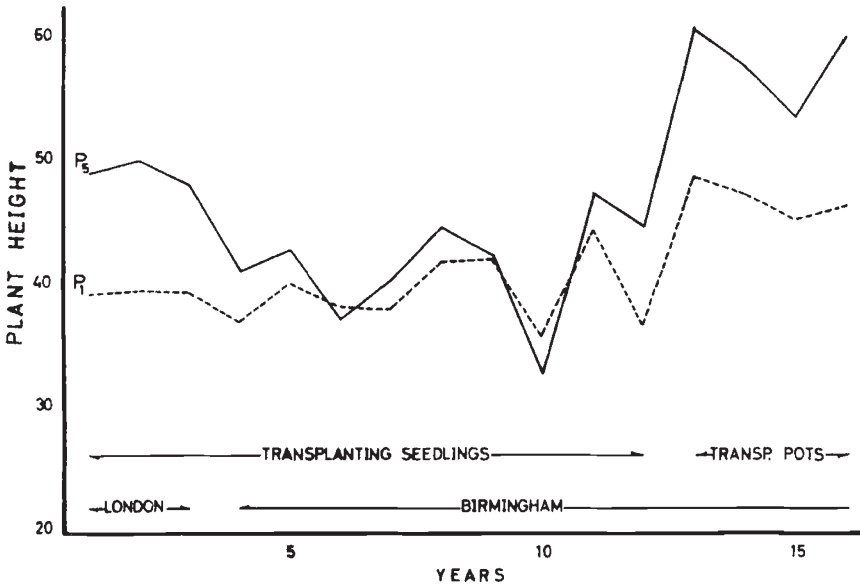

FIG. 1.—Graphical representation of plant height in the inbred lines P_1 and P_5 of *N. rustica* from 1946 to 1964.

In order to have a better understanding of what is meant by stability and performance over environments it is necessary to describe and estimate the environmental and genotype-environmental parameters. For this purpose the model of Mather 1949 and Mather and Morley Jones 1958, as applied to inbred homozygous lines in a series of macro-environments will be used, in which:

$$P_1 = \mu - [d] + \epsilon - \gamma$$
$$P_5 = \mu + [d] + \epsilon + \gamma$$

This means that the phenotypic effect of each line has a genetic effect $[d]$ (following Jinks and Morley Jones's 1958 notation) an environmental effect ϵ and a joint effect of the genotype and the environment γ or genotype-environmental interaction. The genetic value $[d]$ will be considered as a constant and will be taken as the average deviation of the two lines from the over-all mid-parent (μ).

Thus

$$\mu = \frac{\Sigma n_1 \bar{P}_1 + \Sigma n_5 \bar{P}_5}{\Sigma n_1 + \Sigma n_5} \doteq 43\cdot72 \qquad \text{(from table 1).}$$

The environmental effect ϵ is the deviation of the mid-parent value of the two lines in a particular environment from μ, so that

$$\Sigma\epsilon = 0.$$

In this way γ will be the difference between the observed phenotypic value of a given line in a certain environment and its expected value obtained by summing the two additive, genetic and environmental components, so that:

$$\gamma = \bar{P} - ([d] + \epsilon).$$

It should be pointed out that γ will include a random error which cannot be estimated because of the difficulty of replicating a genotype in the same set of micro-environmental conditions. If the model is for macro-environments, then the error will be the within, or the micro-environmental variance.

Therefore if two inbred lines, P_1 and P_5, consistently show the same phenotypic difference under different environmental conditions we can state that these particular genotypes are not interacting with the environment and, moreover, that they have the same stability, *i.e.* $\gamma = 0$; $VP_1 = VP_5$. If, on the other hand, the phenotypic differences between the two lines alters from one environment to another then clearly there is an interaction between the genotype and the environment taking place. The fact that $[d]$ is a constant and ϵ and γ are variables whose values depend upon the particular environment the experiment is grown in, raises two questions.

(a) How to estimate the effects ϵ and γ and
(b) Whether or not ϵ and γ are independent variables.

The answer to the first question can be found by adding and subtracting the phenotypic expressions of $\bar{P}_1$ ($\bar{p}_1$) and $\bar{P}_5$ ($\bar{p}_5$) measured as deviations from μ. These have the following expectations;

$$\epsilon = \tfrac{1}{2}(\bar{p}_1 + \bar{p}_5) \qquad [d] + \gamma = \tfrac{1}{2}(\bar{p}_5 - \bar{p}_1)$$

Table 2 shows the estimated values of ϵ and γ. In this way the observed mean values of the two inbred lines in any environment can be sub-divided into its component parts.

Thus $\quad P_1 \,(1946) = \mu - [d] + \epsilon - \gamma = 43\cdot72 - 2\cdot79 + 0\cdot51 - 2\cdot08$
$$= 39\cdot36$$

and $\quad P_5 \,(1946) = \mu + [d] + \epsilon + \gamma = 43\cdot72 + 2\cdot79 + 0\cdot51 + 2\cdot08$
$$= 49\cdot10.$$

The answer to the second question, namely the possible interdependence of ϵ and γ can be found from the scatter diagram shown in fig. 2 in which the values of ϵ have been plotted against the corresponding values of $[d]+\gamma$. Since $[d]$ is a constant it cannot affect the relationship between ϵ and γ. The reason for its inclusion will become apparent later. From this graph two conclusions emerge; firstly that ϵ and γ are related linearly, that is; the genotype environmental interaction is directly proportional to the environmental effect, and

TABLE 2

Estimated values of ϵ, γ and $[d]+\gamma$

Year	Env.	p_1	p_5	ϵ	γ	$[d]+\gamma$
1946	1	−4·36	5·38	0·51	2·08	4·87
1947	2	−4·12	6·42	1·15	2·48	5·27
1948	3	−4·24	4·44	0·10	1·55	4·34
1950	4	−6·59	−2·55	−4·57	−0·77	2·02
1951	5	−3·52	−0·78	−2·15	−1·42	1·37
1952	6	−5·38	−6·35	−5·87	−3·28	−0·49
1953	7	−5·58	−3·18	−4·38	−1·59	1·20
1954	8	−1·82	0·96	−0·43	−1·40	1·39
1956	9	−1·57	−1·22	−1·40	−2·62	0·18
1957	10	−7·88	−10·86	−9·37	−4·28	−1·49
1958	11	0·73	3·71	2·22	1·30	1·49
1959	12	−6·97	0·90	−3·04	1·15	3·94
1960	13	4·98	17·01	11·00	3·23	6·02
1961	14	3·59	13·97	8·78	2·40	5·19
1962	15	1·51	9·77	5·64	1·34	4·13
1964	16	2·64	16·31	9·48	4·05	6·84

secondly the regression line cuts the y axis at a distance $[d]$ from the origin

so that
$$y = [d]+\beta\epsilon.$$

For the moment this regression line will be referred to as the function of the effect of the environment, and it is in itself a scale which can be used to remove the effect of the genotype-environmental interaction when estimating the average difference of two genotypes in a given environment.

The function of the effect of the environment can be estimated from the calculated values of ϵ and $[d]+\gamma$ in table 2, but it is more desirable to carry out the estimation directly from the absolute phenotypic values of the two inbred lines given in table 1 using the following procedure:

By definition $[d]$ is half the average difference of P_1 and P_5 over all environments.

Then
$$[d] = \tfrac{1}{2}(\bar{P}_5-\bar{P}_1) = 2\cdot79 \qquad \text{(from table 1)}.$$

The required statistics may be found from the fact that

$$VP_1 = V\epsilon + V\gamma - 2 \text{ cov } \epsilon\gamma = 16\cdot05$$
$$VP_5 = V\epsilon + V\gamma + 2 \text{ cov } \epsilon\gamma = 63\cdot10$$
$$V(P_1 + P_5) = 4V\epsilon = 133\cdot56$$
$$V(P_5 - P_1) = 4V\gamma = 23\cdot48.$$

It follows that

$$VP_5 - VP_1 = 4 \text{ cov } \epsilon\gamma \quad \text{and:}$$

$$\beta_{\epsilon\gamma} = \frac{\text{cov } \epsilon\gamma}{V\epsilon} = \frac{VP_5 - VP_1}{V(P_1 + P_5)}.$$

Then

$$V\epsilon = \tfrac{1}{4}V(P_1 + P_5) \qquad \hat{V}\epsilon = 33\cdot39$$
$$V\gamma = \tfrac{1}{4}V(P_5 - P_1) \qquad \hat{V}\gamma = 5\cdot87$$
$$\text{Cov } \epsilon\gamma = \tfrac{1}{4}(VP_5 - VP_1) \qquad \widehat{\text{Cov }} \epsilon\gamma = 11\cdot76$$
$$\beta_{\epsilon\gamma} = \frac{VP_5 - VP_1}{V(P_1 + P_5)} \qquad \widehat{\beta}_{\epsilon\gamma} = 0\cdot35$$

On the left hand side we have the expectations of the parameters we are interested in while on the right hand side are given their estimated values for the *Nicotiana rustica* data. The relative sizes of $V\epsilon$ and $V\gamma$ are of importance to the breeder, but this topic will be discussed at length in a later paper. From the estimated value of β we can write down the equation relating ϵ to γ as:

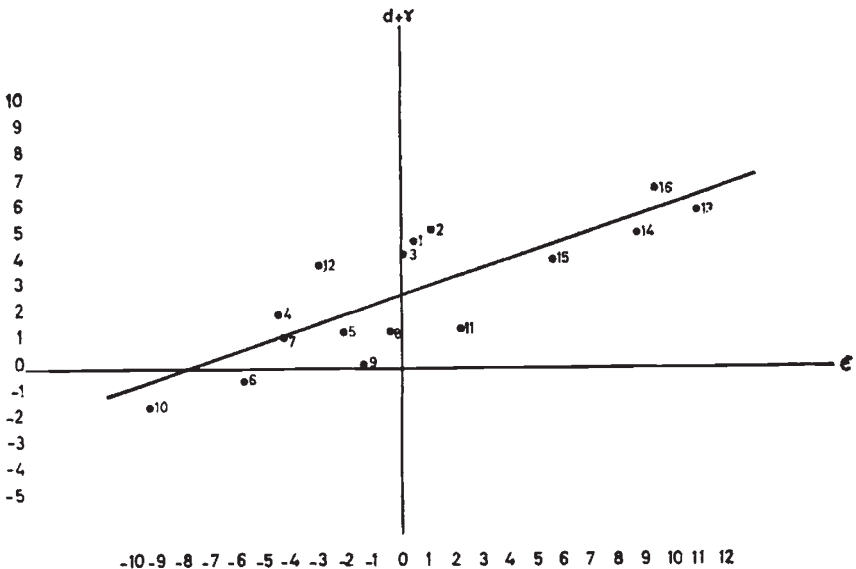$$\nu = 2\cdot79 + 0\cdot35\epsilon.$$

Fig. 2.—The function of the effect of the environment for two homozygous genotypes. Numbers refer to environments in table 2.

This is the regression line shown in fig. 2 and which merits further discussion. Genotypes P_1 and P_5 have been subject to changing environments. Change in locality and management mean that in the present experiment P_1 and P_5 have been subject to a wider range of environmental conditions than would be expected under a random sampling of the environments. The fact that the relationship of ϵ and γ still holds means that $[d]$, the genetic difference, increases with slope β in an improving environment, and it is for this reason that we have referred to the regression line as the function of the effect of the environment.

It is worth noting that the function of ϵ need not necessarily be linear. In fact it is expected that the slope will decline under extreme conditions but in so far as it is linear over the range of environments tested it means that alterations in the environment will result in corresponding changes in the phenotype. Because the function of ϵ is linear, we can recognise several different magnitudes of β, namely:

$$\beta > 1, \quad \beta = 1, \quad \beta < 1, \quad \beta = 0$$

obviously the size of β indicates the values that ϵ and γ have relative to each other, so that when β is greater than 1 the absolute value of γ is larger than ϵ; when $\beta = 1$, $\gamma = \epsilon$ and when $\beta < 1$ the interaction effect γ will be smaller than the environmental effect ϵ. Similarly for those situations in which β is negative. Finally, β can equal zero for one of two reasons. Either there is no genotype-environmental interaction *i.e.* $\gamma = 0$ and all the variation between environments can be ascribed solely to the environmental effect ϵ; or $\gamma \neq 0$ but γ is not a function of ϵ. For example, a situation of this type can be visualised if we compare the phenotypic expression of two inbred lines which are adapted to two entirely different sets of environmental conditions. Furthermore let us suppose that the experiment is conducted in these two environments and all the range of environments in between. The yields of these two hypothetical lines are then expected to change gradually in an inverse way thus:

Environments / Genotypes	E_1	E_2	E_3	E_4	E_5
PA	-2	-1	0	$+1$	$+2$
PB	$+2$	$+1$	0	-1	-2
$\epsilon = \frac{1}{2}(PA + PB)$	0	0	0	0	0

Clearly A and B will have the same variance over environments. On the other hand since A and B are inversely correlated,

$$\epsilon = \tfrac{1}{2}(PA + PB) = 0,$$

2 C

and the function of ϵ does not exist.   This theoretical example shows the classical type of genotype-environmental interaction.   Under such circumstances the line A has to be selected as the best for environments E_4 and E_5 and B for environments E_1 and E_2.   In an intermediate environment such as E_3 however the two lines will have the same performance.

The extreme case described above, although quite frequent, causes no trouble to the breeder for the purpose of selection, since it is easily recognisable.   The more general case is the one in which the genotype-environmental interaction is not very drastic, as in the case of the analyses we have made of P_1 and P_5 in *Nicotiana rustica*, in which

TABLE 3

Phenotypic effects in two homozygous genotypes when empirical values are given to the components [d], ϵ and γ to fit Haldane's situations

Haldane's situations	Empirical values to suit Haldane's situations			ϵ negative		ϵ positive	
	$[d]$	ϵ	γ	$P_A = -[d]-\epsilon+\gamma$	$P_B = [d]-\epsilon-\gamma$	$P_A = -[d]+\epsilon-\gamma$	$P_B = [d]+\epsilon+\gamma$
(1) $[d] > \epsilon > \gamma$	3	2	1	−4	0	−2	6
(2) $[d] > \gamma > \epsilon$	3	1	2	−2	0	−4	6
(3) $\epsilon > [d] > \gamma$	2	3	1	−4	−2	0	6
(4) $\gamma > [d] > \epsilon$	1	3	2	−2	−4	0	6
(5) $\epsilon > \gamma > [d]$	2	1	3	0	−2	−4	6
(6) $\gamma > \epsilon > [d]$	1	2	3	0	−4	−2	6

$\gamma = f(\epsilon)$.   In such cases the relative values of $[d]$, ϵ and γ are important from the standpoint of selection.

Haldane (1946) has described six possible situations in regard to the relative sizes of these components and Mather and Jones (1958) have discussed them.   Table 3 shows the expected phenotypic values which these six situations will have when the numerical values 1, 2, and 3 are given alternatively to $[d]$, ϵ and γ respectively.   Fig. 3 is a graphical representation of the change in phenotypic expression.

The following conclusions can be drawn:

(*a*) When ϵ is positive, *i.e.* when the performance of the two geno-types is better than average, irrespective of the relative sizes of $[d]$, ϵ and γ, P_A will always have greater expression of the character con-cerned than P_B, although the difference $2[d]$ between them may differ.

(*b*) When the performance of P_A and P_B is less than average (ϵ being negative) then the selection of the " best " genotype will depend on the relative sizes of $[d]$, ϵ and γ; for example P_A will be selected if situations (1) (2) or (3) occur, that is:

$$[d] > \epsilon > \gamma$$
$$[d] > \gamma > \epsilon$$
$$\epsilon > [d] > \gamma.$$

FIG. 3.—Graphical representation of expected phenotypic values of two homozygous genotypes in Haldane's situations (from table 3).

But if situations (4) (5) or (6) occur that is:

$$\gamma > [d] > \epsilon$$
$$\epsilon > \gamma > [d]$$
$$\gamma > \epsilon > [d]$$

then P_B will be selected, since it will show a greater phenotypic expression under these particular environmental conditions. These

results do not follow simply because of the particular values we have given to $[d]$, ϵ and γ in table 3, they apply to any set of values which we give to the three parameters. This is because the expectation of $P_5 - P_1$ when ϵ is positive equals $2[d] + 2\gamma$, whilst when ϵ is negative it equals $2[d] - 2\gamma$. It follows therefore that in the former situation this expectation will always be positive, but in the latter it can be positive or negative depending upon the relative magnitudes of $[d]$ and γ. From the standpoint of selection these results are important, since they indicate that those genotypes with a greater expression of the character under consideration are more easily detected when the general performance of the lines is better than average *i.e.* when ϵ is positive, conversely they are less likely to be recognised when ϵ is negative. Clearly in those environments where the relative positions of the lines change frequently, the breeder will have to adapt his selection techniques to fit in with the requirements of a situation which demands that different lines be selected for the different environments as opposed to the situation where one line is suitable for general use.

In the biometrical model which has been adopted for the present investigation of genotype-environmental interaction, the absolute values of $[d]$, ϵ and γ will depend in magnitude upon the particular estimated value of μ for any given set of data.

If the range of environments over which the experiment is conducted can be regarded as a random sample from the population of all possible environments, the estimates of μ, $[d]$, ϵ and γ can likewise be regarded as estimates of the population parameters. On the other hand, if the range of possible environments is extended by virtue of improved techniques then these new values of the parameters will reflect this improvement in the environment. But such an improvement need not necessarily alter the relationship between ϵ and γ, as can be seen from the results of the *Nicotiana rustica* data already described.

Thus it follows that whenever the function of the environment is positive the most desirable genotypes are expected to perform best in the better environment, a fact which has been recognised for some considerable time by both plant and animal breeders.

3. SUMMARY

A model has been developed which will enable the interaction between the genotype and the environment to be investigated in greater detail. This model has been applied to data obtained from two inbred lines of *N. rustica* over a period of 16 years at two locations.

In this particular set of data, analysis of generation means reveals that genotype-environmental interaction is linearly related to the environmental effect. The implications of this result and other possible situations which could arise have been discussed from the practical breeder's point-of-view.

4. REFERENCES

HALDANE, J. B. S. (1946). The interaction of nature and nurture. *Ann. Eugenics*, *13*, 197-205.

JINKS, J. L., AND JONES, R. MORLEY. (1958). Estimation of the components of heterosis. *Genetics*, *43*, 223-234.

JONES R. MORLEY, AND K. MATHER. (1958). Interaction of genotype and environment in continuous variation. II. Analysis. *Biometrics*, *14*, 489-498.

MATHER, K., AND JONES R. MORLEY. (1958). Interaction of genotype and environment in continuous variation. I. Description. *Biometrics*, *14*, 343-359.

MATHER, K. (1949*a*). Biometrical Genetics. Methuen, London.

MATHER, K., AND VINES, A. (1952). The inheritance of height and flowering time in a cross of *Nicotiana rustica*. Quantitative Inheritance. Eds. E. C. Reeve and C. H. Waddington. 49-79 H.M.S.O.