FACTORS OF COVARIATION IN *NICOTIANA RUSTICA*

L. J. EAVES and R. J. BRUMPTON
Department of Genetics, University of Birmingham, England

Received 28.x.71

1. INTRODUCTION

MULTIVARIATE methods have been used widely, though perhaps not always wisely, for investigating genetical variation for complex characters. Studies have been published in plant genetics (*e.g.* Tukey, 1951; Murty, Arunachalam and Jain, 1970), in behaviour genetics (*e.g.* Bock and Vandenberg, 1968; Hegman and Defries, 1970; Royce, Carran and Howarth, 1970) and in animal genetics (Bailey, 1956). Multivariate techniques have also been recognised as potentially effective in the definition of selection criteria in plant breeding (Rao, 1953; Whitehouse, 1969) though this effectiveness has still to be proved.

This study had several aims:

1. To provide a description of the genotypic and environmental covariation for *Nicotiana* as an aid to the selection of characters for further study.
2. To summarise the genetical variation as simply as possible by deriving compound characters, and to estimate the number of genetical factors segregating in a cross between two lines.
3. To compare the information obtained from different multivariate approaches to the analysis of genetical variation.
4. To illustrate techniques which facilitate the interpretation of such studies, especially that of factor rotation.
5. To identify, if possible, groups of similar lines.
6. To determine whether the structure of environmental and genotype-environmental variation was related to the structure of genotypic variation.

Three main procedures have been adopted in the genetical study of multiple characters. There have been various attempts at factor analysis (*e.g.* Murty, Arunachalam and Jain, 1970) to identify traits of genotypic variation. Canonical analysis, or discriminant function analysis has been applied (Mather, 1949; Whitehouse, 1969; Ram and Panwar, 1970) and there have been studies to identify related phenotypes by clustering methods (*e.g.* Somayajulu, Joshi and Murty, 1970).

Factor analysis and canonical analysis are discussed in many texts (*e.g.* Seal, 1964; Morrison, 1967) and Gower (1967) introduces some techniques of cluster analysis.

The difficulties which are encountered when multivariate methods are applied to genetical problems may be summarised as statistical, interpretational or genetical difficulties. Many of the studies reported hitherto have failed to avoid criticism on at least one of these criteria. Whilst this is tolerable when applications for new procedures are being developed, it cannot remain a permanent deficiency of multivariate studies in genetics.

Attempts to apply factor analysis to genetical problems by consideration of genetical correlations derived from variance components have already been criticised on statistical grounds (*e.g.* Seal, 1964) and many of the methods employed for solving the factor model are equally inadequate since they are approximations which do not lead to statistically testable conclusions (Kendall, 1957). The centroid method (*e.g.* Thurstone, 1947) is one such procedure which has assumed a widespread role in behavioural studies and has been adopted in some genetical analyses (Murty, Arunachalam and Jain, 1970). Most algorithms for solving the factor model require a tedious iterative procedure and a principal components solution has often been adopted as an approximation to factor analysis (*e.g.* Bailey, 1956). This assumes, in effect, that all the variance of every variable can be assigned to common factors. This hypothesis should be tested if possible and is likely to be disproved when the correlations between the variables are not large.

The statistical difficulties encountered when analysing matrices of genetical correlations can be avoided by considering the matrix of correlations between line means or the between-line covariance matrix. In this case statistical difficulties are exchanged for interpretational ones, on account of the possible bias due to environmental components of covariation. Large families are needed if the analysis and its interpretation are to be reliable.

The problem of estimating the factor loadings was solved, in principle at least, by Lawley (1940) who derived the equations for the maximum-likelihood estimates of the loadings. Recent developments (Jöreskog, 1967; Clarke, 1970) have made the maximum-likelihood solution more readily attainable, and the precise procedure is discussed in relation to the analysis of covariation for *Papaver dubium* by Gale and Eaves (1972).

The location of reference axes obtained by any factor-algorithm is arbitrary and interpretation is regularly clarified by rotation (*e.g.* Thurstone, 1947). Various criteria have been proposed for analytical rotation to avoid rotation to purely subjective criteria. Thurstone (1947) described the principle of simple structure as a basis for factor rotation and many writers have suggested numerical criteria and computational methods for obtaining the best possible approximation to simple structure for a particular set of common factors.

Kaiser (1958) offered the Varimax criterion which requires that the total within-factor variance of the factor loadings be maximised. Thurstone recognised that simple structure might be obtained more clearly by relaxing the restraint that factors should be orthogonal. Hendrickson and White (1968) suggested the Promax method for oblique rotation which transforms the matrix of factor loadings to a least-squares fit with a derived matrix having a pattern which represents the criteria of simple structure. Few genetical studies, apart from behavioural studies, have employed factor rotation, but the arbitrariness of raw factors and their sensitivity to small changes in the correlation matrix make rotation a feature of factorial studies which is virtually essential. Apart from assisting the interpretation of individual studies, rotation to simple structure may give a better basis for comparisons between studies (Cattell, 1952) than consideration of arbitrary unrotated factors.

Many genetical studies have been based on material of uncertain genetical history (*e.g.* Murty, Arunachalam and Jain, 1970), or on a small sample of inbred lines (*e.g.* Royce, Carran and Howarth, 1970). In the former case

the possibilities for a more detailed genetical analysis are limited and it is difficult to specify the kinds of gene action responsible for the observed covariance structure. In the latter case there can be little confidence that genetical associations between variables reflect anything more than chance associations of genes.

There is thus a need to draw together the most powerful statistical, interpretational and genetical devices into a multivariate study which can provide a basis for future studies. This has been attempted for data from a series of 82 inbred lines derived from the cross of varieties 1 and 5 in *Nicotiana rustica*.

2. Material and methods

The biometrical genetical study of the generations derived from the cross 1 × 5 was started by Mather and Vines (1952). Jinks and Perkins (1970a, b) summarise the earlier studies of this cross and further studies have been reported by Perkins and Jinks (1970, 1971). The 82 inbred lines are the product of a selfing programme initiated by Perkins and Jinks in 1968. One hundred F_2 plants were taken at random and selfed for six generations, two

Table 1

List of characters

	Character	Description
1	FT	Flowering-time in days from 1st June
2	HFT	Height in cm. at flowering time
3	FH	Final height (cm.)
4	H0	Height at 42 days from sowing (cm.)
5	H1	Height at 56 days from sowing (cm.)
6	H2	Height at 70 days from sowing (cm.)
7	H3	Height at 84 days from sowing (cm.)
8	NL	Number of true leaves

every character. Preliminary univariate analyses of variance revealed significant interactions between lines and blocks for some of the variables, but the line means were based on the two blocks jointly since there was no intention of generalising the results beyond these particular lines and blocks. For the purpose of the analyses which follow the data were summarised by the between-lines covariance matrix and the within-lines covariance matrix. The latter was calculated for each line and block separately and the matrices pooled to give the overall within-line dispersion matrix. Except where occasional plants were missing, therefore, each within-line variance or covariance was based on 18 d.f. Data for missing plants, which were few in number, were supplied by including the line mean for the calculation of the between-lines matrix, and by a reduction in the d.f. for the within lines matrix. There was some evidence of heterogeneity among the within-line dispersions which may reflect genotype-environment interaction and could detract from the validity of the canonical analysis to be reported. The covariances and correlations are given for the between-line variation in table 3. The values for the within-line variation are given in table 4. Within-line standard errors were calculated for each line and variable. These were included as additional variables with the line means to give the 30×30 correlation matrix **C**. This matrix is not tabulated because of its size. The between-line covariance and correlation matrices formed the basis of the analysis of the genotypic covariation, the within-line correlation matrix was used to indicate the structure of micro-environmental variation, and the matrix **C** was analysed to reveal the structure of certain genotype-environmental interactions.

(a) *Genetical analysis*

(i) *Factor analysis*

The structure of genotypic variation was examined first by a factor analysis of the between-line correlation matrix. The maximum-likelihood procedure was adopted and the algorithm given by Clarke (1970) was used to secure rapid convergence. The solution required that some specific variances took negative values. Such a solution was termed " improper " by Jöreskog (1967). That more " improper " solutions have not been found reflects the long-standing failure to solve accurately the equations of the maximum-likelihood method. Jöreskog suggested (1967) that the specific variances for the offending variables be set at zero and the maximum-likelihood solution obtained subject to this restriction. This is the procedure which has been adopted here. The extraction of factors was concluded when the chi-square test of goodness of fit (Bartlett, 1954) was reduced to a non-significant value. The correlation matrix was subjected to factor analysis rather than the covariance matrix since it was felt that differences between the raw variances for the different variables can have little biological significance. The number of plants per family is quite large so the between-line variances and covariances will reflect primarily genetic variation.

The raw factors were rotated to simple structure using Varimax to provide an orthogonal solution and Promax for an oblique solution. Second-order factors (Thurstone, 1947) were extracted from the oblique factors to provide an integrated summary of the structure of the genotypic variation. The standard errors of the specific variances were calculated to provide tests of significance of specific genetical variation (Clarke, 1970).

TABLE 3

*Between-line covariances and correlations**

Character														
FT	HFT	FH	H0	H1	H2	H3	NL	D2	D3	D3	LP	LL	LMB	MB
1094·3	2102·4	1875·7	−82·7	−532·4	−1721·1	−1382·4	−46·6	−505·7	−1557·8	−723·1	22·7	384·4	107·5	186·4
88	5266·4	4533·9	−89·1	−726·4	−2608·4	−1864·1	−65·3	−67·8	−1807·8	61·5	290·7	1300·2	507·9	669·4
76	84	5506·1	−110·6	−821·4	−2601·1	−904·5	−52·9	1010·3	−260·5	3374·8	526·4	3200·6	1135·6	2277·2
−64	−31	−38	15·3	73·2	188·6	137·7	5·9	101·7	222·3	95·7	6·8	−47·5	−4·6	−43·0
−80	−49	−55	92	409·8	1095·0	754·0	32·6	355·3	1019·2	323·2	−3·4	−387·0	−108·5	−278·4
−91	−63	−61	84	95	3274·1	2585·7	88·2	1227·4	3099·4	1457·4	37·3	−710·0	−165·4	−421·5
−75	−46	−22	63	67	81	3127·6	73·5	2124·9	3723·6	4026·0	340·1	1161·4	521·3	1072·0
−53	−33	−27	56	61	58	49	7·1	3·7	181·2	128·3	13·6	−1·5	24·6	−22·6
−20	−01	18	34	23	28	50	02	5683·0	4432·9	6206·3	686·5	3615·4	1296·3	2671·8
−48	−25	−03	58	51	55	68	69	60	9641·3	10079·1	934·1	3777·4	2032·4	2129·2
−17	01	35	19	12	20	55	37	63	79	16989·0	1592·2	8680·5	3704·3	5537·7
03	19	33	08	−01	03	28	24	43	44	57	458·2	933·8	649·4	329·9
15	23	56	−16	−25	−16	27	−00	62	50	86	56	5977·4	2288·4	3838·7
09	19	41	−03	−14	−08	25	24	46	55	76	80	79	1415·7	852·9
08	13	42	−15	−19	−10	26	−12	49	30	58	21	68	31	5338·4

* Diagonals and upper triangle give covariances. Lower triangle gives correlations (decimal point omitted).

TABLE 4

Within-line covariances and correlations

Character														
FT	HFT	FH	H0	H1	H2	H3	NL	D2	D3	D4	LP	LL	LMB	MB
15·2	31·5	17·0	−1·4	−6·5	−19·7	−23·0	−1·3	−34·3	−66·9	−59·4	−4·1	−20·3	−9·0	−16·4
70	135·6	97·3	−1·1	−3·2	−7·5	4·9	−1·8	−3·1	−33·6	86·2	12·8	65·4	26·8	63·1
36	68	148·9	−0·5	−1·0	6·7	50·9	−0·9	45·1	37·6	209·0	28·3	138·2	54·8	124·7
−48	−13	−05	0·5	1·5	3·6	3·7	0·3	7·5	13·5	12·5	1·1	4·9	2·2	4·1
−53	−09	−03	64	9·8	18·6	19·0	1·1	36·1	64·2	86·5	5·7	30·1	12·9	27·4
−67	−09	07	65	79	56·9	67·4	2·9	105·1	186·4	231·4	20·1	99·0	43·5	86·4
−53	04	38	46	55	81	121·4	3·4	150·5	254·8	392·0	37·2	185·0	78·5	164·1
−51	−24	−00	53	52	58	46	0·4	5·6	14·1	13·1	1·1	4·8	2·5	3·5
−39	−01	16	46	52	62	61	38	502·3	535·3	613·1	53·8	274·0	113·7	247·1
−52	−09	09	56	62	75	70	65	72	1086·4	1039·0	96·3	456·6	211·7	388·9
−31	15	35	34	56	62	72	40	55	67	2443·2	171·4	978·7	411·7	856·6
−16	17	36	23	28	41	52	27	37	45	54	41·5	105·0	56·8	70·1
−20	22	44	26	37	51	65	28	48	54	77	63	661·8	235·8	485·0
−17	17	32	22	30	41	51	28	37	46	60	63	66	193·0	176·8
−18	23	44	24	38	49	64	23	47	51	75	47	81	55	539·3

(ii) *Canonical analysis*

Whilst the factor analysis of the matrix of between-line correlations gives a model of genetical structure which takes account of both common and specific genetical variation the procedure is deficient in two important respects. Factor analysis, firstly, takes no account of the different heritabilities of the variables in question and, secondly, provides no direct indication of the reliability with which genotypes may be distinguished from one another by the main genotypic factors. The technique of canonical analysis, however, provides the basis for such an investigation. There is no reason why the two approaches should lead to identical pictures directly but the information they provide is complementary and should be reconcilable by rescaling and rotation. Canonical analysis involves calculating the linear compounds of the variables which maximise the discrimination between lines when tested against the variation within lines. The contribution of each canonical variate to the overall between-line discrimination can be assessed by a chi-square test. Transformation of the original measurements provides a set of discriminant function scores with an identity covariance matrix within lines and a diagonal variance-covariance matrix between lines with entries corresponding to the latent roots of $\mathbf{BW}^{-1}$, where **B** is the original between-lines matrix and **W** the original within-lines matrix. The method is potentially powerful for multivariate genetical analysis since any genetical hypothesis which may be tested for the univariate case by analysis of variance may be readily extended to the multivariate case by canonical analysis (*e.g.* Fulker, Wilcock and Broadhurst, 1972).

A canonical analysis was accordingly conducted for these data and the narrow heritabilities calculated for each of the canonical variates. If λ_i is the between-lines variance for the ith variate, that is, the ith root of $\mathbf{BW}^{-1}$, the between-line component of variance is $\frac{1}{n}(\lambda_i - 1)$, since the within-line variance is scaled to unity, where n is the number of plants per family. The narrow heritability is thus $ht_n = \sigma_b^2/(\sigma_b^2+1)$ and provides a guide to the likely rate of response if the discriminant functions were used as a basis for selection (Mather and Jinks, 1971).

The discriminant functions were used to transform the original line means to provide mean scores for each line on the canonical variates. The extreme lines on some of the major variates were used to characterise the dimensions of genotypic variation. Since the average within-line variance, σ^2, is unity, the sum of squares for a comparison between any two lines is a chi-square for 1 d.f. and provides a test of the reliability with which lines can be discriminated by a particular variate.

To elucidate the relationship between the canonical variates and the factors, the transformed means were correlated with the original variables to provide the " loadings " of the original variables on the canonical variates. These were rotated in the same way as the genotypic factors. Second-order factors were extracted as before and were compared with the second-order genotypic factors.

(iii) *Effective factors and grouping of lines*

There has been some speculation in the psychological literature about the relationship between factors and genetical determinants (*e.g.* Royce, 1957;

Thompson, 1966). The availability of so many inbred lines provides an opportunity to estimate the number of effective factors for which the parental lines differ with respect to genotypic factors and thus to obtain the first concrete evidence on the issue. If the additive effects are equal at all loci and the two most extreme lines are represented in this experiment, the number of effective factors, k, for which the parental lines differ is estimated by

$$\hat{k} = \frac{\frac{1}{4}(\bar{L}_1 - \bar{L}_2)^2}{\hat{D}},$$

where $\bar{L}_1$ and $\bar{L}_2$ are the means of the highest and lowest lines respectively, and $\hat{D}$ is the estimate of the additive genetical component, that is $\hat{\sigma}^2_b$ in this case, assuming no non-allelic interaction. The value of $\hat{k}$ will be an underestimate of k if either of the two extremes is not represented or the additive effects of the genes are unequal (Mather and Jinks, 1971).

When we consider the number of effective factors determining variation for correlated characters, two possibilities arise which may be of value. Either we may estimate the number of effective factors shared by pairs of variables or we may estimate the number of effective factors for particular combinations of characters, for example combinations decided on the basis of canonical analysis. There should be some consistency between the results of the two methods.

The number of effective factors common to two variables, designated i and j, may be estimated by

$$\hat{k}_{ij} = r_{ij}\sqrt{(\hat{k}_{ii}\ \hat{k}_{jj})}$$

where r_{ij} is the genotypic correlation between variables i and j, and $\hat{k}_{ii}$ and $\hat{k}_{jj}$ are the numbers of effective factors for each of the same two variables estimated as described above. This is a minimal estimate based on the effective factors having equal and additive effects in both characters apart from differences of scales.

If we write $\mathbf{K}$ for the matrix of common effective factors, $\mathbf{B}$ the between-line covariance matrix, $\mathbf{k}$ the diagonal matrix of $\hat{k}$'s for the original variables, and $\boldsymbol{\delta}$ the diagonal matrix formed by the raw variances of the original matrix, then

$$\mathbf{K} = (\boldsymbol{\delta}\mathbf{k}^{-1})^{-\frac{1}{2}}\mathbf{B}(\boldsymbol{\delta}\mathbf{k}^{-1})^{-\frac{1}{2}}.$$

The between-lines covariance matrix is thus rescaled. Each variance and covariance is now represented as a multiple of the contribution of a single effective factor to the variance of the variables concerned. The magnitudes of these contributions are the elements of the diagonal matrix $\boldsymbol{\delta}\mathbf{k}^{-1}$ corresponding to particular variables.

The number of effective factors for each main canonical variate can be estimated by the usual procedure already described. $\bar{L}_1$ and $\bar{L}_2$ are the two most extreme transformed means on the variate under consideration, and $\hat{D} = \dfrac{1}{n}(\lambda_i - 1)$. With any finite sample of lines the probability of obtaining both extremes decreases as the number of effective factors increases. If the extremes are not represented $\hat{k}$ will be too low. With a sample of 82 lines the extremes are quite likely to be represented, providing fewer than 6 effective factors are involved. Estimates of k which are higher than this can be

regarded as underestimates of the true value. Enough lines were grown in this experiment to permit the distinction between variation due to a few effective factors which might be identified individually from that due to genes too numerous to show clear individual effects.

The problem of considering the number of effective factors may be approached from a further viewpoint by attempting to identify groups of similar lines in the space defined by the canonical variates. If the same genes have equal and additive effects on a number of variables the single factor defined by those variables will generate scores which should fall into a number of groups which will be equal to $k+1$ where k is the number of loci. In this case it is impossible to separate the effects of single genes. If, on the other hand, the genes affect different characters in different directions, or to different degrees, more than one dimension will be required to account for the covariation among the variables and the effects of particular genes might be identifiable if groups are sought in more than one dimension.

It is possible to group lines having similar mean vectors by considering the scores of the lines on the canonical variates. For any pair of lines the sum of squares between lines accumulated over n orthogonal variates provides an indication of the degree of separation of the two lines. When the variates have unit within-line variance, as in the case for the canonical variates, the between-line sum of squares is a chi-square for n d.f. With p lines the variation between lines may be partitioned into $(p-1)$ orthogonal components each with n d.f. Some such sets of comparisons between lines provide groupings for which the within-group variation is small compared with that between groups. This conception forms the basis for grouping the lines.

The procedure adopted in this analysis merely provides a solution which is adequate for descriptive purposes. The method seeks the group which has the largest number of lines within a given criterion distance from the group centre and then identifies successively smaller groups until all the lines are assigned to groups or left isolated. The criterion adopted was the chi-square for n d.f. corresponding to a probability level of 0·01. Such a criterion takes account of different numbers of dimensions by changing the value of n. The analysis only provides a hypothesis for subsequent independent experiments: it is not possible to use the result of such grouping to test the validity of the groups on the same body of data.

For each line in turn a list was prepared of those lines which were within the criterion distance. The largest group was taken as the first cluster of lines. The line nearest the centre of the group was found by selecting that line around which the sum of squared deviations of the other lines is smallest. If the group is large, the number of dimensions small, and the lines distributed uniformly throughout the group, then the line thus selected should be close to the centre of the group. When two groups of the same size overlap the lines falling within the intersection are assigned to the group having the nearest centre. The analysis was conducted for a number of canonical variates both singly and jointly.

(b) *Environmental factors*

The structure of the environmental variation was studied by factor analysis of the within-line correlation matrix. After rotation, the primary orthogonal factors were compared with the primary orthogonal factors of the

genotypic variation. A coefficient of factor similarity was calculated to facilitate this comparison. Where f_{Gi} is the loading of the ith variable on a particular genotypic factor, f_{Ei} the loading of the same variable on the corresponding environmental factor, the coefficient, $C_i = \sum_i f_{Gi} f_{Ei} / \sqrt{\sum_i f_{Gi}^2 \sum_i f_{Ei}^2}$. The structure was summarised by the extraction of second-order factors after oblique rotation of the primary factors.

(c) *Genotype-environment interaction*

Analysis of the **C** matrix of correlations between line means and standard errors provides some indication of the structure of G × E interaction. Genotypic variation in sensitivity to the environment is reflected in the variation from line to line of the within-line standard errors. G × E interaction which is independent of the mean scores of the lines will be indicated by the standard errors loading on different factors from the means. On the other hand, if G × E is determined by the genes which determine between-line variations for the scores, both means and standard errors will load on the same factors (Eaves, 1972). Factors with loadings for means and standard errors having the same sign indicate that a high mean is associated with high variability and low mean with low variability. If means and standard errors load with opposite sign, a high mean is associated with low variability and vice versa.

Since the **C** matrix is partly based on standard errors the assumptions of the factor model are certainly inappropriate so the maximum-likelihood procedure was not adopted. There could be no valid statistical criterion for deciding on the number of common factors. Instead Guttman's criterion was adopted (Guttman, 1954) and as many factors were extracted as corresponded to the eigenvalues of **C** greater than unity. After oblique rotation, second-order factors were extracted, the number again determined by Guttman's criterion. No attempt was made to estimate specific variances in the analysis of G × E interaction and the solution is based on rotation of the principal components of **C**.

3. Results and discussion

(a) *Genetical analysis*

(i) *Factor analysis*

Eight factors were extracted before an adequate fit was obtained ($\chi^2_{(20)} = 14{\cdot}62$, $P > 0{\cdot}05$). This apparently provides little gain in simplicity. Some of the factors are small and could reflect unbroken associations between genes affecting different variables. They might equally result from a few genes which do not show the same pattern of pleiotropic effects as those responsible for the greater part of the variation. The eight varimax factors are given in table 5, with the communalities and specific variances. To obtain a solution of the model seven specifics had to be set to zero. Where the specifics are not zero their standard errors are given, but it is not possible at this stage to give standard errors for the factor loadings (Clarke, 1970). The eight factors account for 90 per cent. of the total standardised variance of the line means for the 15 variables, and for all the significant common variance. The largest loading on the smallest significant factor may be

taken as a rough guide to the lower limit for a significant contribution to the common variation. Such is the loading of 0·14 for H2 on the eighth factor.

The first genotypic factor reflects genetical influences on the development of the plant to flowering-time. The alleles which are responsible for relatively late flowering tend to result in plants which are shorter at the beginning of the season. Thus FT has a loading of 0·61 on this factor whereas H0 and H1 both have loadings of −0·92. The second factor is primarily concerned with the area of individual leaves with loadings of 0·84 on LL and 0·76 on MB. There is clearly some genetical variation for FT, HFT and FH which cannot be attributed to the developmental influences of genes affecting early growth because the third factor is substantially accounted for by these three variables with only small contribution from the earlier measures. Factor 4 is a second leaf-factor, relating more to " stalkiness " than area, since it comprises mainly variation in LL, LMB, and LP. Leaf number forms the basis of factor 5, though this character also has association with the first

TABLE 5

Factor analysis of genotypic covariation (decimal points omitted from loadings)

	Varimax factor										
Character	I	II	III	IV	V	VI	VII	VIII	Communality	Specific (ψ)	s.e. (ψ)
FT	61	−00	75	04	−15	14	04	−05	0·99	0·01	0·002
HFT	22	04	96	13	−09	06	−01	09	1·00	—	—
FH	27	40	78	24	−02	−09	04	−11	0·93	0·07	0·012
H0	−92	−11	−09	00	20	05	13	−07	0·92	0·08	0·013
H1	−92	−14	−27	−09	21	06	−05	−01	1·00	—	—
H2	−86	−04	−44	−03	15	−14	−02	14	1·00	—	—
H3	−70	34	−31	18	13	−50	−02	00	1·00	—	—
NL	−42	−07	−20	18	80	−05	−06	−00	0·89	0·11	0·041
D2	−68	64	−15	35	−10	−02	40	−03	0·81	0·19	0·045
D3	−26	38	−02	31	47	04	32	01	1·00	—	—
D4	06	67	16	46	21	02	03	−03	0·93	0·07	0·011
LP	−01	14	18	78	05	−05	02	−02	0·70	0·30	0·047
LL	01	84	08	54	03	04	−00	04	1·00	—	—
LMB	01	34	05	93	17	01	05	02	1·00	—	—
MB	09	76	11	05	−07	−09	02	−01	0·62	0·38	0·059

factor. Factor 6 is mainly composed of variation in H3 and may reflect the fundamental non-linearity of growth. The remaining two factors are required by the model but they account for very little and we propose no interpretation of the loadings.

The orthogonal solution resolves FT and the successive heights into two factors, 1 and 3 above. The first is predominately developmental and may be explained as the long-term effect of genes which operate early in development, and the second is attributable to genes which are first operative around flowering-time and affect growth subsequent to flowering. The orthogonal model may be too restricting to account for the variation in growth and flowering. It will be shown presently that for purposes of clarity the two factors may be included in one second-order factor affecting rate and duration of development. The number of factors required for leaf characteristics might also be reduced since some of these are shown to contribute to an overall factor of leaf variation.

The communalities of the variables are generally very high which indicates that virtually all the variation for some of the measures is shared with one or more factors. All eight non-zero specifics, though small, are highly

significant, although the standard errors are strictly only applicable to large samples. Specific variances might be due to departures from linearity in the relationship between variables, insipient factors not yet represented properly in the range of measurements, or, indeed, specific genetical influences on particular variables. Only two of the specifics are large, namely LP ($\psi = 0{\cdot}30$) and MB ($\psi = 0{\cdot}38$).

An oblique rotation by Promax was conducted in an attempt to approximate simple structure more closely. When the restraint of orthogonality was removed the developmental factor and the flowering-time factor (factors 1 and 3) were identified as two correlated factors ($r = 0{\cdot}51$) of early growth and flowering-time respectively. The three genotypic factors of leaf variation (2, 4 and 5) were changed little by oblique rotation, but 2 and 4 showed a correlation of 0·67 suggesting that a large proportion of the variation of leaf dimensions can be attributed to the pleiotropic effects of the same genes.

TABLE 6

Comparison of second-order structure for genotypic and environmental factors

Character	Maximum-likelihood genotypic factors		Canonical analysis		Maximum-likelihood environmental factors	
	I	II	I	II	I	II
FT	96	−10	98	02	−16	84
HFT	79	08	81	21	38	73
FH	76	49	73	51	58	51
H0	−66	10	−72	14	32	−54
H1	−83	01	−86	−00	46	−55
H2	−94	10	−96	06	61	−61
H3	−74	56	−78	49	81	−37
DL	−57	10	−60	18	32	−72
D2	−07	71	−51	85	64	−42
D3	−40	53	−26	61	65	−56
D4	−07	75	16	79	88	−20
LP	05	64	13	61	70	−02
LL	27	76	12	93	87	00
LMB	14	64	05	83	70	−03
MB	23	58	18	58	81	01

Factor 5 was now highly specific for leaf number and virtually independent of 2 and 4, although having a correlation of −0·31 with factor 1 suggesting that the genes affecting growth also affect leaf number. In general, plants which flower early tend to have more leaves. The correlation matrix of the eight primary oblique factors was subjected to a components analysis to provide second-order factors. The first two second-order factors accounted for 68 per cent of the variance of the primaries. After varimax rotation to give simple structure for the loadings of the primary factors on the second-order factors the loadings of the original variables on the second-order factors gave the factor pattern in table 6.

This configuration might be approximated more directly by rotation of the two largest primary factors or even principal components (Eysenck and Eysenck, 1968) but this would mean constraining the factor space prematurely in two dimensions and losing the additional information about the relationship between the variables in all eight dimensions. The rotation of only two factors in this case would involve accepting a model which at the outset is known to be inadequate by statistical criteria.

The first second-order factor associates late flowering and comparative tallness at flowering-time and at the end of the season with relative shortness at the start of the season. Late flowering plants appear to start growing later, but continue far longer in development than plants which flower early. Provided all possible genotypes are represented and there are no important chance associations of genes, we may conclude that this factor might be due to the pleiotropic effects of genes which operate early in development yet determine variation for characteristics around flowering-time. Genetical variation which is expressed very early in development is having marked effects throughout the season.

The other factor is seen to be one of leaf development. Some of the variation for leaf length, breadth, plant diameter and petiole length is determined by the same genetical variation. This result can be expected when measurements are all made on a localised part of the plant.

Thus, although the factor model requires eight primary factors to account for the significant common genetical variation the best solution requires that these be oblique. The whole system of primaries can be referred to two second-order axes. This procedure means that it is possible to describe the observed variation with as much generality as possible without losing sight of the primary structure, as reflected in the original factors and specifics.

(ii) *Canonical analysis*

The latent roots of $\mathbf{BW}^{-1}$ are given in table 7 with the chi-square test of the individual components of variance (Hope, 1968). Even the smallest of

TABLE 7

Variances of canonical variates, tests of significance, and narrow heritabilities of compound characters

Variate (i)	λ_i	χ^2	d.f.	ht_n
1	138·48	3238·3	95	0·87
2	38·13	1700·5	93	0·65
3	23·62	1252·3	91	0·53
4	22·62	1216·0	89	0·52
5	18·78	1067·3	87	0·47
6	14·74	893·1	85	0·41
7	12·18	771·5	83	0·36
8	9·75	646·2	81	0·30
9	7·49	519·2	79	0·24
10	4·91	359·9	77	0·16
11	4·39	325·4	75	0·14
12	2·37	184·5	73	0·06
13	2·29	178·5	71	0·06
14	1·88	147·9	69	0·04
15	1·34	106·9	67	0·02

these is statistically significant leading to the conclusion that there is clear distinction between the lines in all 15 dimensions of variation. This is consistent with the finding of the factor analysis that eight of the variables showed significant specific variation and eight common factors were required to account for the common variance. The smaller canonical variates might well reflect significant genetical variation which is specific to particular variables on the factor model. The narrow heritabilities of the canonical variates are also given in table 7. The first eight canonical variates have heritabilities which are clearly larger than some of the original variables and

TABLE 8

Discriminant functions

Character	I	II	III	IV	V	VI	VII	VIII
FT	−0·1172	−0·0874	−0·0983	0·1160	0·1153	0·0413	−0·4375	0·2996
HFT	0·0133	−0·0376	0·0062	−0·0082	−0·0364	0·0738	0·0873	−0·1251
FH	−0·0220	−0·0197	0·0063	−0·0340	0·0339	−0·0998	0·0238	0·0174
H0	−0·1409	−0·6151	−0·5577	0·1837	−0·1752	−0·2412	−0·1196	0·6855
H1	0·0350	−0·1786	−0·0279	0·0099	0·0824	−0·0760	0·0187	0·1153
H2	0·1490	−0·0238	0·0649	0·0768	−0·0100	−0·0559	−0·0858	−0·0229
H3	−0·0006	0·0272	−0·1305	−0·0707	0·0555	0·1099	−0·0761	0·0233
NL	−0·3411	−0·2699	0·8666	−0·7478	0·6241	0·3483	0·4781	−0·2229
D2	−0·0105	0·0068	−0·0247	0·0162	−0·0249	−0·0046	0·0233	−0·0045
D3	−0·0147	0·0005	0·0116	−0·0035	0·0036	−0·0085	−0·0109	0·0009
D4	−0·0034	0·0029	0·0056	−0·0053	−0·0085	−0·0050	−0·0187	0·1153
LP	0·0069	−0·0132	0·0171	0·0064	−0·0331	0·1015	0·1089	0·1214
LL	−0·0063	0·0044	−0·0032	−0·0112	−0·0308	−0·0229	−0·0202	−0·0262
LMB	−0·0013	0·0001	0·0056	−0·0114	−0·0189	0·0013	−0·0078	0·0042
MB	0·0008	0·0149	0·0069	0·0425	0·0514	0·0095	0·0337	0·0119

the first eight latent roots represent 91 per cent. of the trace of $\mathbf{BW}^{-1}$. Since most of the genetical information is represented by these variates, and because eight common factors were required to satisfy the factor model, they form the basis of subsequent interpretation.

The coefficients required to transform the original scores to canonical variate scores for the first eight variates are given in table 8. The raw means were transformed and extreme lines on the canonical variates were considered to characterise the variates. Table 9 gives the raw means for some typical lines.

TABLE 9

Original means for some typical lines

	Line									
Character	17	20	48	74	2	44	18	78	11	70
FT	66·55	62·35	41·15	42·65	64·15	51·25	44·75	44·70	49·55	71·95
HFT	102·78	87·00	54·72	50·55	114·43	61·33	44·00	58·78	57·00	78·86
FH	150·50	141·75	103·85	92·50	156·20	124·00	88·65	114·40	120·95	104·20
H0	1·95	2·13	5·36	3·89	2·78	2·06	2·26	5·12	2·84	1·61
H1	8·55	6·70	27·70	18·15	10·83	9·67	8·55	21·35	12·40	4·50
H2	22·25	22·88	71·58	57·13	33·18	39·38	11·15	59·75	41·60	10·75
H3	60·73	69·08	96·85	85·43	80·58	90·40	44·45	98·93	89·85	31·40
NL	7·70	7·35	9·15	8·40	7·20	7·90	7·92	7·75	8·75	6·50
D2	204·65	191·20	197·10	170·80	191·05	199·75	169·40	214·20	184·10	138·60
D3	198·25	189·65	221·55	183·75	177·80	211·90	207·00	204·00	225·10	120·75
D4	381·50	353·75	352·75	323·75	365·75	383·75	366·50	376·00	414·25	228·75
LP	39·05	30·35	35·85	30·25	42·15	44·10	38·05	37·50	43·50	26·64
LL	222·15	186·31	187·45	176·00	210·00	211·10	201·95	197·65	232·70	135·62
LMB	99·05	85·34	87·80	78·90	97·55	97·80	97·30	90·85	114·15	67·79
MB	159·60	134·25	125·35	134·90	149·20	158·65	132·25	141·80	144·15	98·21
Canonical variate										
1	−16·55	−15·59	−6·15	−7·17	−13·91	−12·42	−9·98	−8·27	−12·37	−15·07
2	−10·88	−9·52	−12·78	−9·80	−12·12	−6·86	−7·22	−11·01	−8·55	−10·87
3	−5·10	−6·25	−5·83	−5·08	−7·57	−6·31	−3·45	−8·39	−5·11	−4·37
4	−2·12	−2·77	−3·14	−2·19	−3·04	−2·39	−3·67	−2·45	−5·39	0·09

Fig. 1 shows the natural logarithms of the heights plotted against time in days from sowing for typical lines chosen to represent the extremes on each of the first two canonical variates. The lines were selected to be about

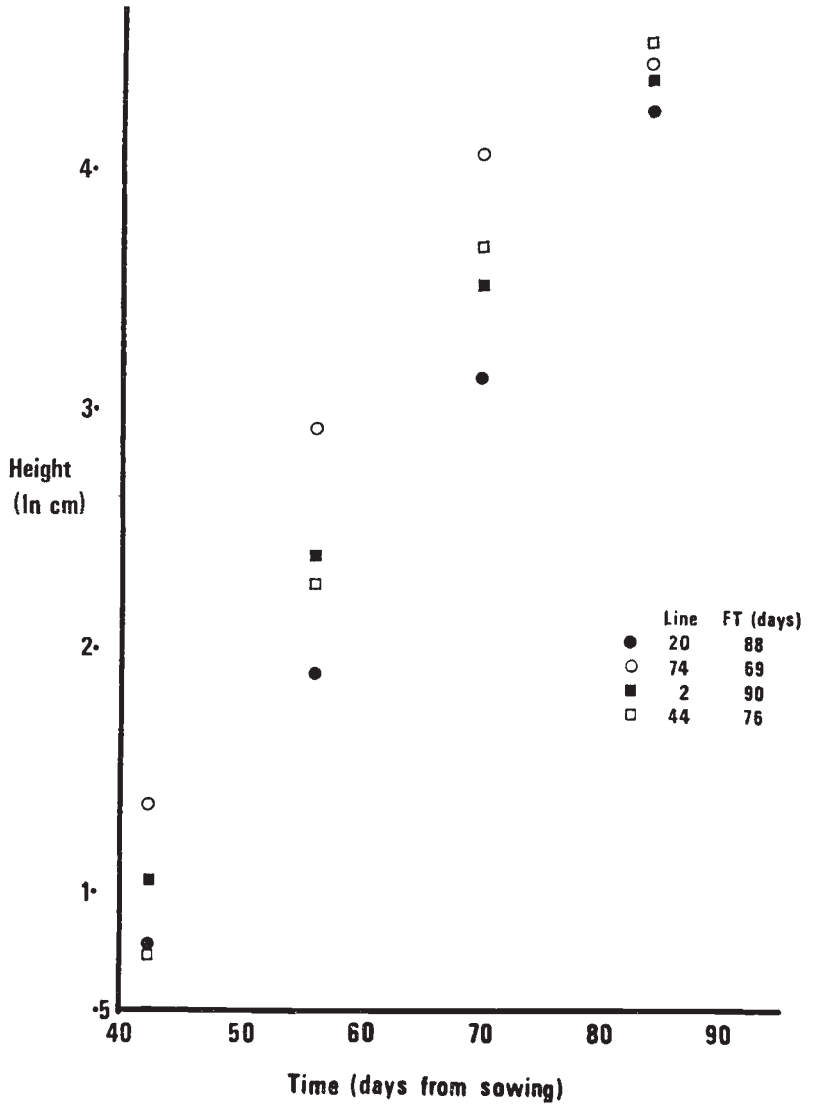

Fig. 1.—The natural logarithm of the height plotted against days from sowing for lines representative of extremes on the first two canonical variates.

average on the other variates. Lines 20 and 74 are the high and low representatives of the first variate, and lines 2 and 44 typify the high and low extremes of the second. The log-growth graphs for lines 20 and 74 are approximately parallel until flowering-time. Growth is much slower after flowering, leading to an intersection of the two curves since the high-scoring

line flowers earlier and stops growing sooner than the low-scoring line. If we approximate the growth until flowering by a logarithmic function

$$h = Ke^{at}$$

where h is the raw height, K is H0, a is a growth constant, and t is the time from H0, then for the first canonical variate, discrimination between the lines is based on K and flowering-time, since a must be approximately the same to give parallel graphs. Lines differentiated on this axis thus have similar growth rates between H0 and FT, but in the high-scoring lines FT occurs earlier, and the plants are taller at H0 than in the low-scoring lines.

The graphs for lines 2 and 44, on the other hand, intersect before flowering-time. In this case the extreme lines are differentiated once more by flowering-time, and H0, although FT may not play a major role in this dimension. It is now the low-scoring line which is taller when the height is first measured but which flowers later, in contrast to the high-scoring line which is shorter at first but flowers earlier. The lines differ again in the value of K, but differences in the growth parameter, a, permit the shorter plants to develop more rapidly.

The two-factor structure of variation for growth revealed both here and in the primary factor analysis may be explained if it is assumed that a certain height threshold has to be reached before flowering can occur, and that in different situations the species may be confronted with different combinations of conditions early in development and at flowering-time.

The possession of a genetical system which can readily achieve early or late flowering whilst being short or tall early in the season may reflect the fact that the species has been exposed to a variety of early and late environments in its natural situation. An annual volunteer habit of *Nicotiana rustica* in the wild (Goodspeed, 1954) is consistent with this interpretation. An integrated system, in which variation in early height was inseparably associated with variation in flowering-time may reflect a consistent association between early and late environments. The second-order factor structure, which links tallness at the start of the season with early flowering suggests that, for the lines in this study, both growth rate and early height form a relatively integrated system related to variation in flowering-time but the presence of two primary factors suggests that this association can be readily broken by recombination.

Lines 17 and 48 are the most extreme lines on the first canonical variate, having scores of $-16{\cdot}56$ and $-6{\cdot}15$. A difference of only 0·62 between the means is significant at the 5 per cent. level so these extremes are highly distinct ($\chi^2_{(1)} = 1083{\cdot}68$). When these lines were grown in the subsequent year the differences between them were obvious to a casual observer.

Line 48 is also the high extreme of the second canonical variate, but line 2 is the next most extreme, and is intermediate on the first variate. Line 40 represents the low extreme, and the difference is again marked ($\chi^2_{(1)} = 344{\cdot}57$). The genetical variation accounts for 65 per cent. of the total variation of this variate.

The third variate is represented at the extremes by lines 18 and 78 which are differentiated mainly by leaf measurements, notably D2, although H3 makes a consistent contribution to discrimination in this dimension. Low-scoring lines tend to have smaller leaves and are somewhat shorter later in the season, but there is no discrimination on the basis of flowering-time.

The two extremes of the fourth variate are typified by lines 11 and 70, although there is also some discrimination between these particular lines on the first dimension, which is seen in the earlier flowering of line 11 compared with line 70, leaf characteristics again form the basis of discrimination on this variate. The leaves of line 70 are shorter, narrower, and rather fewer in number than those of line 11. As an approximation we can represent the area as the product of leaf length and breadth, the length being measured as the difference between total length and petiole length. To this approximation, the leaves of line 11 have an area two and a half times as great as those of line 70 at the time of measurement. If the number of leaves is also taken into account the total leaf area of the two lines may differ by a factor of $3\frac{1}{2}$. There is no overall association between leaf area and flowering-time, although there is apparently an association between large area and early flowering for the two particular inbred lines chosen to represent the extremes.

TABLE 10

Correlations between raw line means and canonical variates (decimal point omitted)

	Canonical variate							
Variable	I	II	III	IV	V	VI	VII	VIII
FT	−92	−25	−01	19	03	13	−10	−06
HFT	−74	−55	−19	05	04	14	08	−23
FH	−80	−26	−34	−23	18	−13	17	−10
H0	72	−49	−20	−11	−05	−20	09	19
H1	88	−36	−05	−10	01	−20	11	08
H2	94	−11	−13	−18	00	−11	10	01
H3	65	11	−48	−49	13	−06	16	01
NL	45	−24	38	−61	16	−07	18	10
D2	06	07	−64	−06	−32	−32	41	04
D3	28	−05	−11	−57	−07	−33	25	12
D4	−09	14	−27	−59	−08	−37	32	−01
LP	−16	−07	−17	−48	−34	−23	54	35
LL	−40	26	−35	−40	−17	−32	34	−11
LMB	−30	08	−12	−60	−38	−04	33	12
MB	−26	36	−37	06	41	−23	48	−06

As the relative contribution of the canonical variates to overall discrimination between lines becomes less it is much more difficult to characterise the extremes. Some impression of the basis for discrimination may, however, be obtained from considering the loadings of the original variables on the canonical variates, which are given in table 10. The smaller variates are seen to reflect further discrimination on the basis of leaf characteristics. The overall structure revealed by the second-order factors extracted from this table of loadings (table 6) are in general agreement with those obtained by rotation of the maximum-likelihood factors. The sole exception is the loading for D2, which is low on the first second-order factor of the canonical variates compared with the loading obtained by factor analysis.

A general conclusion from this part of the study is that factor analysis and canonical analysis lead to substantially similar conclusions with regard to an overall picture of genotypic covariation, but factor analysis provides an indication of the relative importance of specific effects, whereas canonical analysis facilitates the comparison of particular lines.

(iii) *Effective factors and grouping of lines*

Table 11 gives the estimates of the numbers of effective factors involved for each variable individually and jointly, obtained as described above. Inspection of the leading diagonal of the matrix in table 11 shows that about four to five effective factors are involved in flowering-time, final height, and height at flowering-time, six to eight in early heights and leaf number, and seven to ten for plant diameter and leaf measurements with the exception of LMB. Between two and four of the factors involved in the determination of early heights are also implicated in variation at and around flowering-time. About four effective factors could have common influences on all the leaf characters measured. The first row of this matrix provides a model of the way in which genes operating early in development influence later characteristics of the plant. Of an estimated five effective factors which are implicated in the variation for flowering-time, about three are evident when the first height, H0, is measured. By the time H2 and H3 are measured virtually all the genotypic differences which influence flowering-time can be discerned. Some of the factors which are expressed early on do not contribute to later variation. Thus about three of the effective factors influencing H0 are not involved in flowering although they do appear to influence early growth. Such a conclusion is consistent with the view that a separate mechanism is at least partly involved in the determination of early variation, and that flowering and early growth do not form a system which is integrated in any permanent way in *Nicotiana*.

TABLE 11

Estimated numbers of effective factors (sign omitted from off-diagonals)

(a) *Original variables*

	FT	HFT	FH	H0	H1	H2	H3	NL	D2	D3	D4	LP	LL	LMB	MB
FT	5	4	3	3	4	5	5	3	1	3	1	1	1	0	0
HFT	4	5	4	2	3	3	3	2	0	1	0	2	1	1	1
FH	3	4	4	2	3	3	1	2	1	0	2	4	2	2	2
H0	3	2	2	6	6	5	4	4	2	4	4	1	0	0	1
H1	4	3	3	6	7	6	5	4	2	3	1	2	1	0	1
H2	5	3	3	5	6	6	5	4	2	3	1	1	1	0	1
H3	5	3	1	4	5	5	8	4	4	5	5	2	2	2	2
NL	3	2	2	4	4	4	4	7	0	5	3	0	2	2	1
D2	1	0	1	2	2	2	4	0	9	5	6	6	4	3	4
D3	3	1	0	4	3	3	5	5	5	7	7	4	4	1	2
D4	1	0	2	4	1	1	5	3	6	7	10	8	7	4	5
LP	1	2	4	1	2	1	2	0	6	4	8	9	7	4	6
LL	1	1	2	0	1	1	2	2	4	4	7	7	8	5	3
LMB	0	1	2	0	0	0	2	2	3	1	4	4	5	5	1
MB	0	1	2	1	1	1	2	1	4	2	5	6	3	1	8

(b) *Canonical variates*

	Variate							
	1	2	3	4	5	6	7	8
k	3·94	5·83	4·01	9·17	4·37	7·51	4·30	7·75

The values of $\hat{k}$ obtained for the first eight canonical variates are also tabulated and are largely consistent with the pattern of loadings in table 10 and the estimates obtained for the individual variables. The first variate, for example, gives a value for $\hat{k}$ of 3·94 and loads mainly on measurements of

height and flowering-time. The second reflects variation in height independently of flowering-time and yields $\hat{k} = 5{\cdot}38$ and is probably related to the genetic system affecting variation in growth rate and early development. For the fourth variate, $\hat{k}$ is 9·17 and the high loadings are mainly for leaf characteristics. Estimates of the number of effective factors for complex characters are likely to be more reliable than those for simple characters since the identification of the extreme lines is subject to less error than would be the case for single characters. The estimates obtained are also consistent with the interpretation of the canonical variates already offered, since they appear to reflect the number of effective factors thought to be in common for particular groups of variables.

TABLE 12

Grouping of lines on first canonical variate

Score on canonical variate	Line number
−6	48
	24 79
−7	74
−8	7 22 35 38 78
	9 10 26 30 41 66
−9	1
	18
−10	12 21 36 40 77
	19 34 43 53 55 56
−11	29 31 33 49
	16 46 57
−12	11 27 44 52 59 69
	25 73 80
−13	39 47 58 62 63
	2 4 51 60 64
−14	3 5 14 28 45 61
	6 13 50 54 67 72
−15	23 68 70 71 81
	15 20 65 82
−16	8 37 75
	17
−17	

The values obtained for the numbers of effective factors suggest that sufficient loci are involved in the variation to make the identification of individual gene effects impracticable. Most conceivable circumstances lead to the number of effective factors being an underestimate of the number of loci (Mather and Jinks, 1971). In particular, it is possible that linkages are only partly broken during selfing and that extreme lines may not be represented in the sample if the true number of genes is large. Even with as few as nine effective factors, which seem to be involved in differences in leaf measurements, it is unlikely that both the extremes are represented. All these values must, therefore, be treated as minimal estimates of the number of loci.

The results of grouping the lines in the space defined by the canonical variates is summarised for the projections of the mean vectors on to one, four and eight dimensions. In the first dimension the grouping of the lines s quite marked (table 12) and only three lines cannot be assigned to any

group. Grouping of the lines in other single dimensions was less marked since the discrimination effected by the remaining variates, though significant, is not so striking. The groups which are reported, therefore, are to some extent arbitrary divisions of the space in which the distribution of the lines is virtually continuous.

As the number of dimensions is increased the size of the groups is, naturally, reduced and the number of lines which cannot be assigned to any groups rises sharply. In four dimensions 52 lines are assigned to 18 groups, leaving 30 lines unassigned. When eight dimensions are considered, 63 lines cannot be placed in any group, whilst the remaining 19 lines are classified into 8 groups. Thus, by considering the eight discriminant functions based on the mean vectors of 15 measurements for 20 plants per line, it is possible to distinguish virtually every line individually. The number of identifiable genotypes is clearly large, although this does not necessarily suggest that the number of genes is large. Whilst the higher-order factors and some of the canonical variates may reflect the influences of independent sets of genes, some of the discrimination between lines on different variates might be attributed to the unequal contribution of individual genes to the variations of different measurements.

In the system under consideration there appear to be two main sets of genes, one involving four to six effective factors influencing growth and development, and the other involving about nine effective factors influencing leaf formation. The biometrical-genetical analysis, therefore, indicates that the two parental lines, which, it will be remembered had similar heights and flowering-times, differ at a minimum of 13-14 loci with effects on the 15 measurements considered here. In the space defined by the eight canonical variates there are 71 distinguishable genotypes. Even if all the differences can be attributed to the same loci having different patterns of influence on all the variables then at least 7 genes must be involved to enable 71 genotypes to be identified. This gives a logical minimum to the number of effective factors segregating from the original cross. That the number is greater than this is evident from the biometrical analysis which gives estimates of $k > 1$ for all the canonical variates and from the grouping of lines since there are clearly more than two groups in each of the main dimensions of genotypic variation.

The results of attempting to group the lines are consistent with the biometrical analysis in requiring a large number of effective factors to account for the observed variation. The most plausible synthesis requires two independent sets of genes which are represented by the two second-order factors of genotypic variation. Both these factors account for the correlated variation due to a number of genes with effects on growth and leaf formation which could be pleiotropic.

(b) *Environmental factors*

Eight primary orthogonal factors were extracted from the environmental correlation matrix. These were inadequate as a model for the observed covariation ($\chi^2_{(18)} = 126{\cdot}08$, $P > 0{\cdot}001$) but the residual covariation was negligible on inspection and failure of the model may be attributed to the large sample size. Some of the factors seemed trivial after rotation. The eight Varimax factors are given in table 13 and should be compared with the

genotypic factors in table 5. The similarity between the rotated genotypic and environmental factors is very clear and is confirmed by the large values for the indices of factor similarity for the first six factors. Since the factor pattern for the environmental primaries is so similar to that of the genotypic primaries it is unnecessary to provide an individual interpretation of the environmental factors. We must conclude from this similarity that the environmental variation distinguishes the factors that are determined by the genotypic variation. This implies that the micro-environmental influences affect the organism through the same physiological systems as genotypic differences. Cattell (1965) postulated " environmental-mold traits " of behaviour which would reflect the structuring of environmental influences independently of the genotype. There is some evidence for this phenomenon

TABLE 13

Factors of environmental variation and coefficients of similarity to genotypic factors (decimal points omitted)

	Factor										
Character	I	II	III	IV	V	VI	VII	VIII	Communality	ψ	S.E. (ψ)
FT	56	−14	69	−09	−12	19	−00	10	0·88	0·12	0·005
HFT	03	10	99	09	−08	04	01	03	1·00	—	—
FH	11	37	64	23	−03	−30	00	−09	0·70	0·30	0·012
H0	−65	07	−10	09	21	02	−03	−16	0·52	0·48	0·019
H1	−77	21	−09	10	12	11	15	−12	0·73	0·27	0·013
H2	−87	28	−09	22	11	−15	02	−16	0·96	0·04	0·006
H3	−48	54	−02	30	14	−29	27	−25	1·00	—	—
NL	−46	09	−18	14	85	−01	04	−09	1·00	—	—
D2	−42	25	−01	17	08	−02	04	−85	1·00	—	—
D3	−56	31	−09	27	32	−06	14	−39	0·77	0·23	0·009
D4	−36	67	07	31	13	−02	52	−17	1·00	—	—
LP	−16	26	09	78	08	−05	04	−11	0·74	0·26	0·051
LL	−21	75	12	48	07	00	00	−13	0·87	0·13	0·028
LMB	−17	41	09	60	09	00	07	−09	0·59	0·41	0·0025
MB	−22	92	14	23	04	−03	−01	−15	0·83	0·17	0·046
Similarity coefficient	91	90	94	94	96	−81	18	09			

in relation to extraversion (Eaves, 1970), but the present study suggests that the structure of environmental influences is in some cases a clear reflection of the structure of genetical covariation. It should be recognised from the procedural point of view that the comparison of structure would have been less convincing without the refinement of rotation to simple structure.

The eight orthogonal primaries were further rotated to oblique simple structure. Two second-order factors extracted which are given in table 6 and should be compared with the second-order factors of genotypic covariation. At this level there is a substantial general factor of environmental variation with high loadings on every variable except flowering-time. This is in contrast to the second-order structure of genotypic variation where leaf and height variation are virtually independent. The other second-order factor reflects the environmental influences on development until flowering, and is parallel to the structure found for genetical variation. The orthogonal approximation to simple structure at the primary level shows that the same factors may be identified in the genetical and environmental variation. It is clear, however, from the second-order structure, that the organisation of

the primary factors is somewhat different for the genetical variation when this is compared with the environmental variation. The difference may be attributed to the fact that the environmental influences on leaf characteristics in this cross also exert some influence on growth, giving rise to the positive loadings of the successive heights on the first second-order factor. The difference should not obscure the overall similarity of structure between the genetical and environmental variation. The fact that the height measurements load on both second-order environmental factors accounts for their rather smaller loadings on the growth factor, compared with their loadings on the genotypic factor of growth.

The specific variances are predictably larger in proportion to the common variance for the environmental variation (table 13) than for the genetical variation (table 5). This is due, in part at least, to uncorrelated errors of measurement. This is confirmed by the fact that the proportion of specific variance declines with successive height measurements since the earlier measurements are proportionally less reliable than those made when the plants are taller.

(c) *Genotype-environment interaction*

Seven latent roots of **C** were greater than unity and the factors corresponding to these were at once rotated and three second-order factors extracted from the correlated primaries, this being the number required to satisfy the criterion that factors should correspond to roots greater than unity. The loadings of means and standard errors on the three second-order factors are given in table 14.

The growth of the plant until flowering once more forms the basis of the first factor, but the important fact to emerge from this analysis is that the means and standard errors load on the same factor with the same sign, though the loadings for the standard errors are rather lower than those for the means. Genotypes which tend to be tall at the end of the season and flower late also tend to be more variable at the end of the season, but the same genotypes are also shorter and less variable at the start of the season. The suggestion follows that the genes which determine variation in the rate and duration of growth also determine the responsiveness of the organism to differences in the environment. Since the taller plants have a longer period of growth, and these are the plants which are more variable, it might follow that this greater variability results from the longer period of physiological activity in which state the plant is more sensitive to environmental differences. There is, however, no evidence from the cross of varieties 2 and 12 of *N. rustica* that lines selected for high and low sensitivity display correlated patterns of growth in the early stages. Cooper (1969) reports that the late flowering cultivar of *Lolium perenne*, S23, shows less variation in heading date from year to year than early-flowering cultivars which are in active growth during a critical period early in the season.

The second factor shows that much of the variation for sensitivity is common to all characters and is determined independently of genotypic differences in mean performance because the loadings of the line means on this factor are all negligible. The third factor is the genotypic factor of leaf formation which is now shown to be independent of genotypic variation in responsiveness to micro-environmental differences.

For the characters included in this study, therefore, the genes determining sensitivity to micro-environmental differences confer similar covariance structure to those determining a line's mean performance over the range of environments. For leaf characteristics the two aspects are, in part, under independent genetical control, but for characteristics of growth and flowering both sensitivity and mean performance are under the control of related

TABLE 14

Second-order structure of genotype-environment interaction

		Factor		
	Character	I	II	III
Means	FT	91	12	01
	HFT	69	18	18
	FH	66	14	50
	H0	−80	−04	09
	H1	−91	−04	−03
	H2	−95	−07	05
	H3	−75	−08	47
	NL	−65	−20	22
	D2	−21	06	74
	D3	−56	−13	72
	D4	−15	−07	92
	LP	08	−03	63
	LL	24	02	90
	LMB	18	−11	79
	MB	10	05	69
Standard errors	FT	50	46	−25
	HFT	62	41	17
	FH	21	70	23
	H0	−44	34	−08
	H1	−52	44	−27
	H2	−19	75	−08
	H3	41	81	01
	NL	20	47	−27
	D2	41	53	−05
	D3	50	64	−04
	D4	44	64	−02
	LP	23	56	−09
	LL	15	66	13
	LMB	03	35	−18
	MB	05	65	21

genes. In the light of the demonstration of factors of G × E interaction the analysis of the within-line covariance structure must entail a measure of uncertainty because at least some of the within-line variation must be genotype-environmental rather than purely environmental in origin.

4. SUMMARY AND CONCLUSION

1. Multivariate techniques were used to investigate the structure of genotypic, environmental, and G × E covariation for 15 variables in 82 unselected inbred lines derived from 100 F_2 plants from the cross between varieties 1 and 5 in *Nicotiana rustica*.

2. Eight primary genotypic factors were extracted but these could be summarised by two second-order factors, one related to growth and flowering

time, the other to leaf morphology. The factors are independent and probably reflect independent polygenic influences.

3. A canonical analysis gave results which were consistent with the factor analysis but gave additional direct information about the narrow heritability of complex characters, and permitted the estimation of the number of loci segregating. Roughly four effective factors were involved in differences for flowering-time and growth, and about nine with differences in leaf formation.

4. An attempt to group the lines on the basis of their discriminant function scores showed that 71 lines were unique in the space defined by the eight most significant canonical variates.

5. The factors of environmental covariation, with which some G × E will be confounded, were remarkably similar to the genetical factors, suggesting that micro-environmental differences distinguish structures already implicit in genetical covariation. Two second-order factors were extracted from the eight orthogonal primaries. These were: (i) a general factor of environmental variation; (ii) a factor of environmental variation in leaf morphology.

6. The G × E variation for successive heights and flowering-time was shown to be correlated with the genotypic variation for these characters. There was also a general factor of G × E variation which was independent of the genotypic variation in mean performance.

It remains to be seen whether multivariate methods have any long-term genetical value, but this study attempts to elucidate points of procedure and to show how data on large samples with many lines, in conjunction with more powerful analytical techniques can lead to greater consistency and clarity in the interpretation of trait covariation.

Acknowledgments.—The authors are indebted to Professor J. L. Jinks and Dr Jean M. Perkins for their helpful advice and comments on the manuscript. The first author was supported from a grant by the Medical Research Council for research in psychogenetics, and the second author by a Ministry of Agriculture, Food and Fisheries research studentship. The computations were performed on the Birmingham University KDF-9 Computer and the factor rotations were conducted with a programme kindly loaned by the University School of Education.

5. References

BAILEY, D. W. 1956. A comparison of genetic and environmental principal components of morphogenesis in mice. *Growth, 20*, 63-74.

BARTLETT, M. S. 1954. A note on the multiplying factors for various chi-squared approximations. *J. Roy. Statist. Soc. B., 16*, 296-298.

BOCK, R. D., AND VANDENBERG, S. G. 1968. Components of heritable variation in mental test scores. In *Progress in Human Behavior Genetics*, pp. 233-260 (ed. S. G. Vandenberg). Johns Hopkins, New York.

CATTELL, R. B. 1952. *Factor Analysis: an Introduction and Manual for the Psychologist and Social Scientist.* Harper and Bros., New York.

CATTELL, R. B. 1965. Methodological and conceptual advances in evaluating hereditary and environmental influences and their interaction. In *Methods and Goals in Human Behavior Genetics*, pp. 95-140 (ed. S. G. Vandenberg). Academic Press, New York.

CLARKE, M. R. B. 1970. A rapidly convergent method for maximum-likelihood factor analysis. *Brit. J. math. statist. Psychol., 23*, 43-52.

COOPER, J. P. 1969. Developmental genetics. In *Jubilee Report, 1919-1969.* Aberystwyth; Welsh Plant Breeding Station.

EAVES, L. J. 1970. Aspects of Human Psychogenetics. University of Birmingham; Unpublished Ph.D. thesis.

EAVES, L. J. 1972. The multivariate analysis of certain genotype-environment interactions. *Behavior Genetics, 2*, 241-244.

EYSENCK, H. J., AND EYSENCK, S. B. G. 1968. A factorial study of psychoticism as a dimension of personality. *Multiv. behav. Res. special issue*, 15-31.

FULKER, D. W., WILCOCK, J., AND BROADHURST, P. L. 1972. Studies in genotype-environment interactions. I. Methodology and preliminary multivariate analysis of a diallel cross of eight strains of rat. *Behavior Genetics, 2*, 261-287.

GALE, J. S., AND EAVES, L. J. 1972. Variation in wild populations of *Papaver dubium*. V. The application of factor analysis to the study of variation. *Heredity, 28*, 135-149,

GOODSPEED, T. H. 1954. The genus *Nicotiana*. *Chronica botanica, 16*. Waltham, Massachusetts.

GOWER, J. C. 1967. A comparison between some methods of cluster analysis. *Biometrics, 23*, 623-637.

GUTTMAN, L. 1954. Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149-161.

HEGMAN, J. P., AND DEFRIES, J. C. 1970. Maximum variance linear combinations from phenotypic, genetic and environmental covariance matrices. *Multiv. behav. Res., 5*, 9-18.

HENDRICKSON, A. E., AND WHITE, P. O. 1964. Promax: a quick method for rotation to oblique simple structure. *Brit. J. statist. Psychol., 17*, 65-70.

HOPE, K. 1968. *Methods of Multivariate Analysis*. London University Press, London.

JINKS, J. L., AND PERKINS, JEAN M. 1970*a*. A general method for the detection of additive, dominance, and epistatic components of variation. III. F_2 and backcross populations. *Heredity, 25*, 419-429.

JINKS, J. L., AND PERKINS, JEAN M. 1970*b*. Environmental and genotype-environmental components of variability. VII. Simultaneous predictions across environments and generations. *Heredity, 25*, 475-480.

JÖRESKOG, K. G. 1967. Some contributions to maximum-likelihood factor analysis. *Psychometrika, 32*, 443-482.

KAISER, H. F. 1958. The Varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

KENDALL, M. G. 1957. *A Course in Multivariate Analysis*. Griffin, London.

LAWLEY, D. N. 1940. The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edin. ser A., 60*, 64-82.

MATHER, K. 1949. *Biometrical Genetics*. Methuen, London.

MATHER, K., AND JINKS, J. L. 1971. *Biometrical Genetics*. Chapman-Hall, London.

MATHER, K., AND VINES, A. 1952. The inheritance of height and flowering time in a cross of *Nicotiana rustica*. In *Quantitative inheritance*, pp. 401-480 (eds. E. C. Reeve and C. H. Waddington). H.M.S.O., London.

MORRISON, D. F. 1967. *Multivariate statistical methods*. McGraw-Hill, New York.

MURTY, B. R., ARUNACHALAM, V., AND JAIN, O. P. 1970. Factor analysis in relation to breeding system. *Genetics, 41*, 179-189.

PERKINS, JEAN M., AND JINKS, J. L. 1970. Detection and estimation of genotype-environment interaction, linkage and epistatic components of variation for a metrical trait. *Heredity, 25*, 157-177.

PERKINS, JEAN M., AND JINKS, J. L. 1971. Analysis of genotype × environment interaction in triple test cross data. *Heredity, 26*, 203-211.

RAM, J., AND PANWAR, D. V. S. 1970. Intraspecific divergence in rice. *Ind. J. Genet. and Plant Breeding, 30*, 1-11.

RAO, C. R. 1953. Discriminant functions for genetic differentiation and selection. *Sankhya, 12*, 229-246.

ROYCE, J. R. 1957. Factor theory and genetics. *Educ. and psychol. Meas., 17*, 361-376.

ROYCE, J. R., CARRAN, A., AND HOWARTH, E. 1970. Factor analysis of emotionality in ten inbred strains of mice. *Multiv. behav. res., 5*, 19-48.

SEAL, H. 1964. *Multivariate Statistical Analysis for Biologists*. Methuen, London.

SOMAYAJULU, P. L. N., JOSHI, A. B., AND MURTY, B. R. 1970. Genetic divergence in wheat. *Ind. J. Genet. and Plant Breeding, 30*, 47-58.

THOMPSON, W. R. 1966. Multivariate experiments in behavior genetics. In *Handbook of Multivariate Experimental Psychology*, pp. 711-731 (ed. R. B. Cattell). Rand McNally, Chicago.

THURSTONE, L. L. 1947. *Multiple Factor Analysis*. University of Chicago Press, Chicago.

TUKEY, J. W. 1951. Components in regression. *Biometrics, 7*, 33-69.

WHITEHOUSE, R. N. H. 1969. Canonical analysis as an aid in plant breeding. 2nd Barley Genetics Symposium.