ALTERNATIVE WAYS OF ESTIMATING THE NUMBER OF GENES IN A POLYGENIC SYSTEM BY GENOTYPE ASSAY

PHILOMENA TOWEY and J. L. JINKS

Department of Genetics, University of Birmingham, Birmingham B15 2TT, England

SUMMARY

The procedures for estimating k, the number of genes, or more strictly the number of effective factors in a polygenic system by the method of genotype assay have been extended to any number, p, of F_{n+2} grand progeny families raised from each F_n individual assayed. Formulae are also derived that would be more appropriate for estimating k if dominance were absent or in the more unlikely event of no internal balancing.

The existing and new procedures are illustrated by the analysis of data from a cross between varieties 1 and 5 of *Nicotiana rustica* which extended to the F_8 generation. The structure of these data permitted the estimation of k for flowering time and final height in the F_2 from assessments made on the F_4 for $p = 2$, from F_5 to F_8 for $p = 4$ and from joint consideration of the F_4 to F_8 assessments. All these assays of the F_2 gave low estimates for k. On the other hand, using the same F_5 to F_8 data to assay the F_3 to F_6 generations respectively for $p = 2$ gave estimates of k that increased rapidly with generation so that for every one detected in the F_2 there were ten on average in the F_6.

Checks and controls of the material and method, including using the same procedures to estimate k for a known single major gene difference segregating in this cross, leave no doubt that the rise in the estimate is genuine. Furthermore, it is expected from the nature of effective factors and the linkage disequilibrium that is generated on making this cross and subsequently resolved over successive rounds of recombination.

These analyses confirm the overwhelming superiority of genotype assay over the other methods of estimating k that are available in all but a few species.

1. INTRODUCTION

JINKS AND TOWEY (1976) described a new approach to estimating k, the number of genes, or more correctly effective factors in a polygenic system using genotype assay. Basically, each individual from a random sample of m individuals of the F_n generation is assayed for evidence of heterozygosity through two of its randomly chosen grand-progeny families of the F_{n+2} generation. This, however, is but one special case of a general procedure in which the m F_n individuals are each assayed through p grand-progeny families each consisting of l sibs. Within the same total number of mpl individuals in the F_{n+2} generation there are many ways of deploying resources that will have consequences for the reliability of the estimate of k. In this paper we consider some of the theoretical consequences of varying p while keeping m and l constant and also the genetical situations in which intermediate probabilities are more appropriate than P_{Max} for estimating k.

The procedures are illustrated by the analysis of *Nicotiana rustica* breeding programmes based upon the cross V1 × V5 (Mather and Vines, 1952). This analysis provides overwhelming confirmation that the number of

effective factors increases over the successive generations derivable from an initial cross.

2. THEORY

(i) *Setting the limits*

Jinks and Towey (1976) gave the probability, $P_{Het.\,r}$, that a heterozygote in the nth generation would be heterozygous at r of the k loci, where r could take all values from 1 to k as

$$P_{Het.\,r} = \frac{1}{(2n-1)}\, k^{\,k}C_p(2^{n-1}-1)^{k-r}.$$

This expression depends on n, the generation, and k only. However, the probability of detecting differences between the p individuals chosen at random from the F_{n+1} progeny of a selfed heterozygote in the F_n generation is dependent on p and in our earlier paper we considered only the special case of $p = 2$. Where each genotype has a unique phenotype the probability of detecting differences among p individuals will be related to r, the number of loci at which the grand-parent was heterozygous, as

$$1 - \left(\frac{2^p+2}{4^p}\right)^r.$$

From which it follows that the frequency of heterozygotes (P_{Max}) in the F_n generation that is detectable by progeny testing p F_{n+2} random progenies of each individual in the F_n generation will be

$$P_{Max} = \frac{1}{(2^{n-1})^k} \sum_{r=0}^{k} {}^{k}C_r(2^{n-1}-1)^{k-r}\left(1 - \frac{2^p+2}{4^p}\right)^r,$$

which simplifies to

$$P_{Max} = 1 - \left(1 + \frac{2^p+2-4^p}{2^{n+2p-1}}\right)^k.$$

If, following Jinks and Towey, the effects of both internal and relational balance are taken into account the probability that p individuals chosen at random from the F_{n+1} progeny of a selfed heterozygote will give progenies that differ becomes

$$1 - \frac{1}{4^{pk}} \sum_{r=0}^{k} \left(\frac{3^r k!}{r!(k-r)!}\right)^p$$

and the frequency of heterozygosity in the nth generation that will be detectable by our procedures is then,

$$P_{Min} = 1 - \left(1 - \frac{1}{2^{n-1}}\right)^k \sum_{r=0}^{k} \frac{{}^{k}C_r \sum_{s=0}^{r} 3^{ps}({}^{r}C_s)^p}{[2^{n+2p-1}-2^{2p}]^r}.$$

In fig. 1 are plotted the values of P_{Max} and P_{Min} for $k = 1$ to 20, $n = 2$ and 5 and $p = 2$ and 4. For lower values of k both P_{Max} and P_{Min} are more sensitive to changes in k for $p = 4$ than for $p = 2$. Futhermore, the differences between the P_{Max} and P_{Min} curves are less for $p = 4$ than for
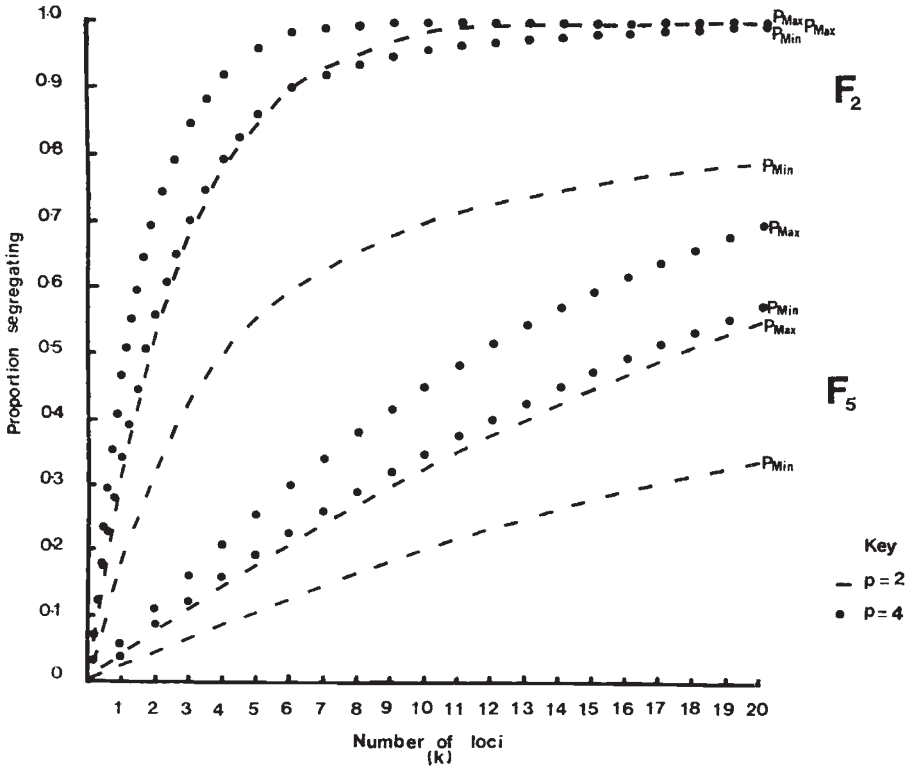
Fig. 1.—The effect of varying p, the number of F_{n+1} individuals chosen from a selfed individual in generation n, on the relationship between the proportion of detectable segregations and k, the number of loci for the maximum and minimum curves. Two values of p and n are used to illustrate the effect.

$p = 2$, and hence the estimates are less sensitive to the assumptions which distinguish them.

(ii) *Intermediate situations*

P_{Min}, which sets the upper limit to the number of effective factors, is based upon the supposition that both internal balance and dominance are operating simultaneously to minimise the number of different genotypes which have different phenotypes. Where there is prior knowledge of the dominance relationships, alternative formulae may be more appropriate. And while it is very unlikely that prior knowledge of the extent of internal balancing would be available it is illuminating to consider the consequences of modifying this assumption.

If there is little or no dominance, so that the phenotype of the heterozygote for any locus is distinct from that of either homozygote but there is maximum internal balance, *i.e.* $d_a = d_b = d_c \ldots = d_k$ the P_{Min} becomes

$$P_{Int. A} = 1 - \left(1 - \frac{1}{2^{n-1}}\right)^k \sum_{r=0}^{k} \frac{{}^kC_r \sum_{s=0}^{2r} ({}^{2r}C_s)^p}{\left[2^{n+2p-1} - 2^{2p}\right]^r}.$$

The alternative situation, namely, high dominance ($h = 2d$, not $h = d$ as stated by Jinks and Towey, 1976) but no internal balance can also be specified by modifying the P_{Min} to become

$$P_{Int.\,B} = 1 - \left(1 + \frac{3^p + 1 - 4^p}{2^{n+2p-1}}\right)^k.$$

But it is doubtful whether in practice there could ever be a situation in which there is no internal balance unless k is very small. For example, for $k = 2$ the condition for no internal balance is simply $d_a \neq d_b$ while for $k = 3$ the conditions required are $d_a \neq d_b \neq d_c$, $d_a \neq d_b + d_c$, $\pm d_a \neq d_b - d_c$, $d_b \neq d_a + d_c$, $\pm d_b \neq d_a - d_c$, $d_c \neq d_a + d_b$ and $\pm d_c \neq d_a - d_c$. For $k = 4$ the number of conditions grows correspondingly since they include in addition to $d_a \neq d_b \neq d_c \neq d_d$ all possible combinations of the types $d_a \neq d_b + d_c + d_d$, $d_a \neq d_b - d_c + d_d$ and $d_a \neq d_b - d_c - d_d$; $d_a + d_b \neq d_c + d_d$ and $d_a + d_b \neq d_c - d_d$. Clearly, as k increases the number of conditions that have to be simultaneously met to ensure no internal balancing becomes impossibly large. In practice, therefore, some internal balancing must always be occurring even when $d_a \neq d_b \neq d_c \ldots \neq d_k$ unless k is very small. The true upper limit to the estimate of k must therefore be somewhere between that obtained from P_{Min} and $P_{Int.\,B}$ but probably closer to that from P_{Min}.

The use of P_{Min} or $P_{Int.\,B}$ to set the upper limit to the value of k strictly applies only if the detection of heterozygotes in the F_n depends solely on the finding of differences among the means of the p F_{n+2} families derived from each F_n grandparent. They are not the appropriate limits, however, if we also detect heterozygotes by finding differences among the variances within the p F_{n+2} families. Thus while dominance reduces the probability of observing differences between a homozygous dominant and a heterozygote on the basis of their phenotypic contributions to family means there is no corresponding reduction arising from their contribution to the within family variances. We should not use variances therefore, in conjunction with the P_{Min} or $P_{Int.\,B}$ formulae. In practice the issue does not arise because we rarely if ever detect a difference between family variances without also detecting a difference in the family means because of the greater sensitivity of the latter. We frequently, however, detect differences between family means without finding a difference between the corresponding family variances (see Jinks and Towey, 1976, Table 2).

In fig. 2 we present $P_{Int.\,A}$ and $P_{Int.\,B}$ for $n = 2$ and 5, $k = 1$ to 20 and $p = 2$ along with P_{Max} and P_{Min} for comparison. Although $P_{Int.\,A}$ and $P_{Int.\,B}$ give probabilities which fall between those of P_{Max} and P_{Min}, $P_{Int.\,B}$ is almost the same as P_{Min} and deviates from it only as n decreases and k increases. $P_{Int.\,A}$, on the other hand, is more like P_{Max} but again it falls progressively below this value as n decreases and k increases.

3. MATERIALS AND METHOD

Appropriate material for illustrating the use of the probability curves for estimating k and the effect of varying p against a constant total of pml individual plants in the generation of assessment is provided by the cross between V1 and V5 of *Nicotiana rustica* initiated in 1944 by Professor K. Mather and his colleagues to study the variation in flowering time and

final height (see Mather and Vines, 1952; Breese, 1954). The structure of the experiment which continued to the F_8 generation grown in 1952 is shown in fig. 3 taken from Breese (1954). Beyond the F_2 only one of the 20 groups of families is illustrated, a group being all the descendants of one of the 20 F_2 individuals. Each group is divisible into two sub-groups, the members of each being the descendants of a single F_3 individual. From the F_5 onwards the experiment was designed to have 20 such groups each consisting of two sub-groups and each sub-group consisting of two families, although it fell below this in the later generations because of random losses
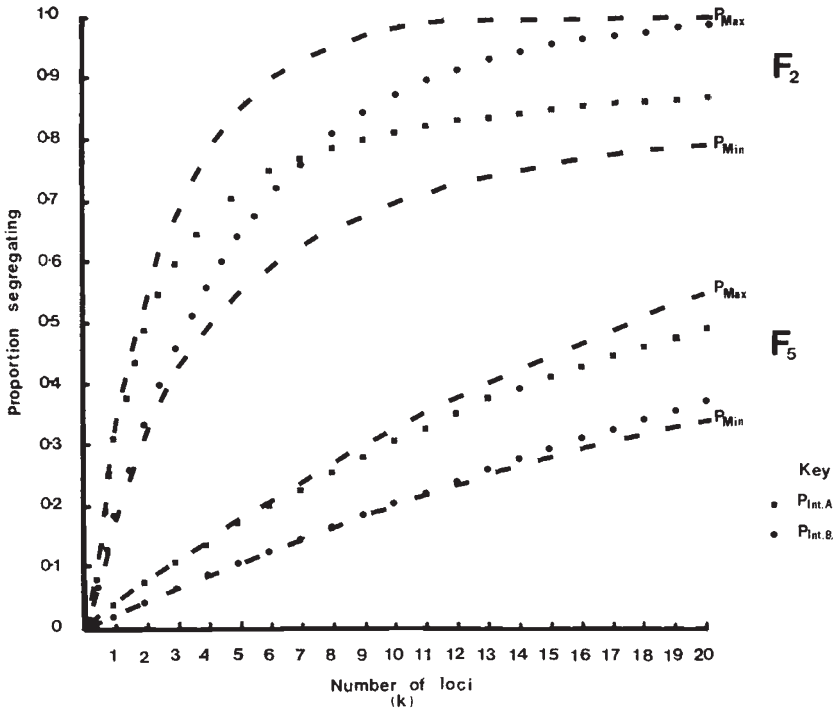

FIG. 2.—The relationship between the proportion of detectable segregations and k, the number of loci for the maximum and minimum curves, and for the two intermediates for $p = 2$. The F_2 and F_5 generations are again used for illustration (see fig. 1).

(see table 1). The unit of randomisation throughout was a plot of five plants of the same family and one plot per family was raised in each of the two independently randomised blocks.

The generations up to the F_4 were grown at Merton, London, and the F_5 to F_8 at Winterbourne, Birmingham. Because of practical difficulties which arose during the transition, the F_5 generation which provided the parents of the F_6 were raised in 1949 (see fig. 3) but the F_5 data we shall place most reliance upon for assaying the heterozygosity in the F_3 were a replicate sample of the F_5 families grown along with the F_6 in 1950.

We can estimate k for the F_2 to F_6 generations by analysing their F_4 to F_8 grand-progeny families for $p = 2$. Since, however, from the F_5 to F_8 generations each F_2 individual is represented by a group of four families

we can use the F_5 to F_8 generations with $p = 4$ to detect segregation and hence estimate k for the F_2, that is, use the F_{n+2} to F_{n+5} generations to estimate k for constant $F_n = F_2$ and constant pml apart from random losses for $p = 4$. Hence the only variable that could affect the estimates of k
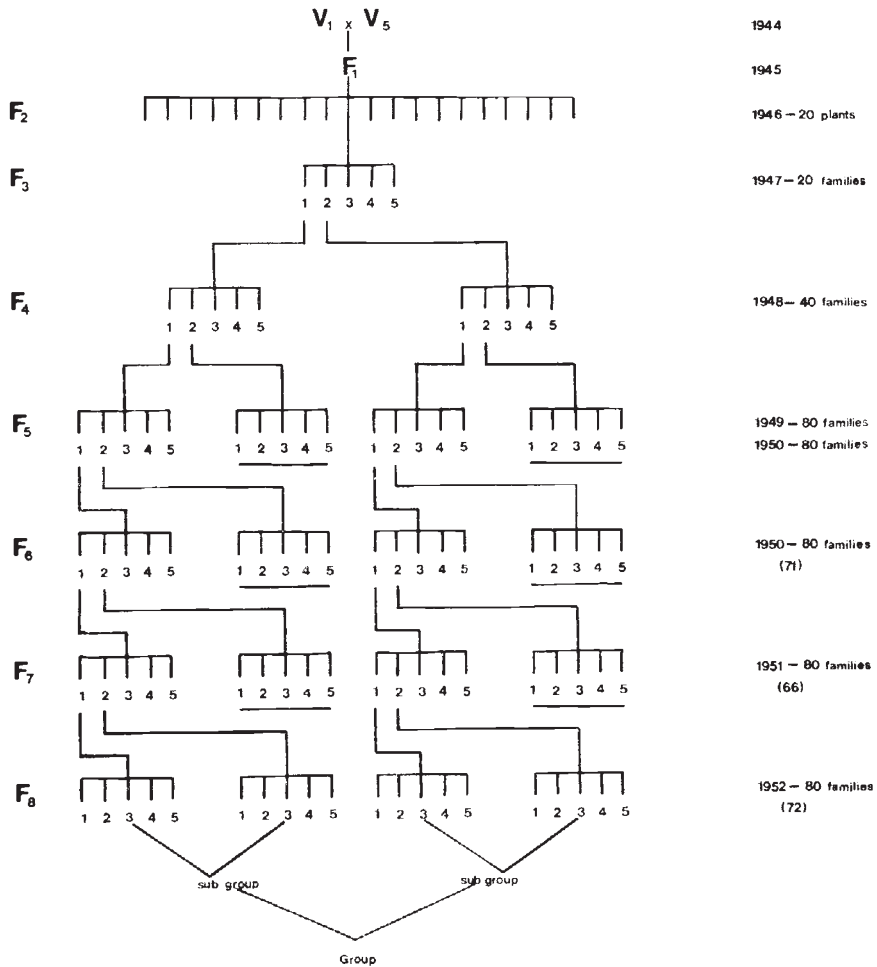

FIG. 3.—Genealogical table showing how a single group of families is derived from an F_2 individual. Each group consists of two sub-groups each derived from an F_3 individual. Twenty such groups each consisting of two sub-groups were initiated in the early generations. Each family is represented by a plot of five plants. The potential number of families are given for each generation together with the actual number achieved (Breese, 1954).

would be any changes in the sensitivity with which we can detect differences between family means from the F_5 to the F_8 generations. We would, for example, expect an increase in sensitivity because of the expected decrease in the variances of the family means as the families become more inbred. But superimposed upon this there is the unpredictable effects of seasonal

differences over the three years, 1950, 1951 and 1952, in which they were grown.

For the estimates based upon $p = 2$, each of the pairs of families within a sub-group was subjected to an analysis of variance in which four items were recognised; the difference between the means of the two families for 1 degree of freedom, the difference between the two blocks for 1 degree of freedom, the interaction between families and blocks for 1 degree of freedom and the differences between individuals within families within blocks ($=$ within plots) for 16 degrees of freedom. If the block interaction mean square was not significant when tested against the within family within

TABLE 1

Summary of the stages in the estimation of k, *the number of effective factors, for flowering time and final height in the* F_2 *to* F_6 *generations for* p $= 2$. *Estimates for* k *in the* F_3 *based upon the* F_5 *raised in 1949 are given in brackets*

Character	Flowering time					Final height				
F_n generation: m in F_{n+2} generation:	F_2 20	F_3 40	F_4 35	F_5 30	F_6 36	F_2 20	F_3 40	F_4 35	F_5 30	F_6 36
P for $\leqq 0.05$	0·200	0·450 (0·200)	0·314	0·233	0·306	0·400	0·300 (0·200)	0·257	0·200	0·306
for $\leqq 0.01$	0·200	0·250 (0·125)	0·086	0·166	0·166	0·250	0·175 (0·125)	0·174	0·166	0·250
k from P_{Max} for $\leqq 0.05$	1	4(2)	5	7	19	2	3(2)	4	6	19
for $\leqq 0.01$	1	2(1)	2	5	10	1	2(1)	3	5	15
k from P_{Min} for $\leqq 0.05$	2	8(3)	9	12	>20	3	4(3)	7	10	>20
for $\leqq 0.01$	2	4(2)	2	8	16	2	2(2)	5	8	>20
for F_{n+2} generation E_2	2·58	4·55	4·55	7·87	3·31	3·83	13·12	13·12	7·00	4·29
σ_ω^2	7·52	22·39	14·86	9·62	9·92	13·82	27·27	20·95	12·10	10·56
Smallest significant difference	3·30	2·40 (3·63)	2·50	2·50	2·00	4·10	3·98 (4·32)	4·60	3·80	2·58

blocks mean square the latter was used to test the difference between the two families. If, however, the block interaction was itself significant it was used to test the difference between the two families, leading to a variance ratio for one and one degree of freedom. This made it practically impossible to find a difference between the families whenever the block interaction was significant.

For the estimates based upon $p = 4$ each set of four families within a group was subjected to an analysis of variance in which the same four items were recognised and the same comparisons were made but with 3, 1, 3 and 32 degrees of freedom, respectively.

In addition to flowering time and final height, one further character was recorded by Mather and Vines and Breese, namely, the presence or absence of anthocyanin on the ovary. This is controlled by a single major gene difference *A-a*, with the presence of anthocyanin dominant to its absence (Mather and Vines, 1952). We can, therefore, use this character as a

control for testing both the material and the method. We can use its segregation in the F_{n+2} to detect heterozygosity in the F_n generation and to estimate the proportion of heterozygotes for both $p = 2$ and $p = 4$. Furthermore, we can base our estimate solely on phenotype differences, *i.e.* where we do not distinguish AA and Aa, in which case the P_{Min} formula for estimating k is appropriate or alternatively use evidence of subsequent segregation to distinguish Aa from AA, in which case the appropriate formula for estimating k is P_{Max}.

4. ESTIMATES

In table 1 are presented the critical steps in the estimation of k, the number of effective factors, for flowering time and final height in the F_2 to F_6 generations, based upon comparisons of their F_4 to F_8 grand-progenies, respectively, for $p = 2$. For example, in the F_5 grown in 1950, 40 pairs of families, each pair having the same F_3 grandparent, were raised. For flowering time 18 of these 40 pairs had significantly different means with a probability of 5 per cent or less (≤ 0.05), and ten at a probability of 1 per cent or less (≤ 0.001). Thus the proportion P of F_3 plants which were demonstrably heterozygous was 0·450 at the 5 per cent level of significance and 0·250 at the 1 per cent level. If we assume that all genotypic differences lead to phenotypic differences, these proportions are estimates of P_{Max} which in turn takes the values expected for $k = 4$ and 2, respectively. If, on the other hand, we assume that a proportion of the genotypic differences are not displayed as phenotypic differences consistent with complete dominance and equal additive genetic effects at all loci, these proportions are estimates of P_{Min} and these estimates are as expected for $k = 8$ and 4 respectively. The other estimates in table 1 have been similarly computed. Since in the seasons in which the assessments were made the V1 × V5 cross displayed dominance for early flowering and for greater final height (Breese, 1954; Bucio Alanis, Perkins and Jinks, 1969), P_{Min} would seem to be more appropriate for estimating k than either P_{Max} or P_{Int}.

In table 2 are presented the corresponding stages in the estimation of k for flowering time and final height in the F_2 based upon assessments of F_5 to F_8 progenies for $p = 4$. Comparisons of the estimates in tables 1 and 2 present some remarkable features. The estimates of k in the F_2 are low whether estimated from the F_4 for $p = 2$ (table 1) or from the F_5 to F_8 for $p = 4$ (table 2). In complete contrast, the estimates of k in the F_3 to F_6 generations based on F_5 to F_8, respectively for $p = 2$, show a marked and progressive increase with generation (table 1) even though they are based on the same observations as the F_2 estimates in table 2. There are two possible causes of any change in k with generation, one of which is common to both sets of estimates and the other unique to the set in table 1. The common cause is changes in the sensitivity with which differences between family means can be detected in the F_{n+2} generation of assessment, *i.e.* F_4 to F_8. Statistics relating to this cause are summarised in the bottom three rows of table 1 but they also relate to the corresponding F_{n+2} columns of table 2.

The first of these statistics, E_2, measures the variation among family means that is expected to arise solely from non-heritable causes (Mather and Jinks, 1971). Any difference between family means must be significantly

greater than this before we can infer heritable differences. The estimates of E_2 which are taken from Mather and Vines (1952) and Breese (1954) vary markedly between the F_4 and F_8 generations, but major trends are discernible. For flowering time there is an increase from the F_4 to F_7 followed by a sharp drop to the F_8. These changes are presumably related to seasonal differences. For final height there is an abrupt increase between the F_4 and the F_5 coinciding with the move from Merton to Winterbourne (see Section 3), followed by a steady decline as the new site was developed. These changes presumably reflect the level of homogeneity of the soil conditions.

The second of these statistics σ_ω^2 measures the average within family variation arising from all causes both heritable and non-heritable in the F_{n+2} generation. It is a component of the sampling error of the family means and hence contributes to the sensitivity with which we can detect

TABLE 2

Summary of the stages in the estimation of k, *the number of effective factors for flowering time and final height in the F_2 using the F_5 to F_8 generations for* p = 4. *Estimates based on the F_5 raised in 1949 are given in brackets*

Character	Flowering time				Final height			
Generation of assessment:	F_5	F_6	F_7	F_8	F_5	F_6	F_7	F_8
m:	20	15	13	16	20	15	13	16
P for $\leqq 0.05$	0·850	0·866	0·769	0·875	0·550	0·666	0·615	0·563
	(0·500)				(0·600)			
for $\leqq 0.01$	0·700	0·400	0·769	0·313	0·500	0·266	0·462	0·563
	(0·100)				(0·300)			
k from P_{Max}								
for $\leqq 0.05$	4(2)	4	3	4	2(2)	2	2	2
for $\leqq 0.01$	2(1)	1	3	1	2(1)	1	1	2
k from P_{Min}								
for $\leqq 0.05$	5(2)	6	4	6	2(3)	3	3	3
for $\leqq 0.01$	3(1)	2	4	1	2(1)	1	2	3

differences between them. Indeed, in the absence of block interactions it is the sole component of the error variance used to detect segregation.

The estimates in table 1 are again taken from Mather and Vines (1952) and Breese (1954). For both characters they show an abrupt increase between the F_4 and F_5 followed by a gradual decrease. The heritable component of this variance, of course, halves with every generation of selfing and this contributes to the steady fall from the F_5 to F_8. The large increase between F_4 and F_5 presumably results from the change in location noted earlier.

The cumulative effects of these changes is revealed by the third statistic in table 1, which is the smallest difference between any pair of family means which proved to be significant at the 5 per cent level ($\leqq 0.05$). For flowering time this decreases from the F_4 to the F_8 but for final height it reaches its greatest value in the F_6 before falling to a relatively low value by the F_8.

If we examine the estimates of k in table 2, for which F_n is constant but the generation of assessment is changing, we find no trends in the estimates for either character at the 1 per cent or the 5 per cent levels of

significance. Indeed, at the latter level the estimates are remarkably stable over the four generations used for assessment. There is no evidence, therefore, that changes in sensitivity are markedly affecting these estimates of k.

Since the same F_5 to F_8 data are also used to obtain the estimates of k in table 1 we do not expect any marked trends in these estimates arising from changes in sensitivity. The large and consistent increases in the estimates with generation shown by both characters at both levels of significance must therefore arise from other causes. Since the estimates in table 2 differ from those in table 1 solely in that F_n is increasing from F_2 to F_6, the cause of this rapid upward trend in the estimates can only be due to this. For different F_n's we use different curves relating P_{Max} and P_{Min} to k and the relative accuracies of the estimates in different parts of the range of k values varies with n (Jinks and Towey, 1976; and figs. 1 and 2). This could not, however, account for the ten-fold or greater increases in estimates from the F_2 to the F_6. Most of this increase must, therefore, be genuine and result from the successive rounds of recombination that occur during the production of the F_2, F_3, F_4 and F_5 gametes.

As a final check on the material and method we have estimates k for the single major gene controlled presence and absence of anthocyanin (Section 3). For example, if we use the 20 pairs of F_4 families ($p = 2$) to determine the detectable heterozygosity for this character in the F_2 on the basis of phenotypic differences only, the proportion of heterozygotes is 0·10 (2 out of 20). Equating this to P_{Min} gives an estimate of $k = 1$. If on the basis of subsequent segregation we distinguish heterozygotes from the dominant homozygote the proportion of detectable heterozygosity in the F_2 rises to 0·40 (8 out of 20). This is now an estimate of P_{Max} and as such it gives an estimate of $k = 2$, the actual value falling between 1 and 2. If for example we now use the alternative estimate of k in the F_2 based upon $p = 4$ in the F_8 and again confine ourselves to phenotypic differences the proportion of detectable heterozygotes is 0·25, which being an estimate of P_{Min} gives $k = 1$. If, on the other hand, we now use the same F_8 data to estimate k in the F_6 we obtain an estimate of $k = 0$ for P_{Min} although the subsequent segregation of a heterozygote would give a P_{Max} estimate of $k = 1$. These estimates give no reason for doubting either the reliability of the material or the validity of this method of analysis.

5. Pedigree analysis

The pedigree of every F_3 to F_8 family can be traced back to one of 20 F_2 individuals (fig. 3), hence we can combine the assessments already made on the F_4 to F_8 generations to individually assay each of the F_2 individuals for heterozygosity. For each F_2 we have five separate occasions from the F_4 to F_8 to determine whether progenies derived from it are segregating, and evidence of segregation at any one of these five occasions is sufficient to establish that the F_2 plant was a heterozygote.

In table 3 are given the distributions among the 20 F_2 groups of the number of occasions for which there is evidence of segregation at the 5 and 1 per cent levels of significance. The proportion showing one or more significances can be regarded as estimates of P_{Max} or P_{Min} for $p = 4$ since there are four families per generation for all but the F_4 for each F_2 group. These estimates equate to the values of k given in table 3. For the 5 per cent signi-

TABLE 3

The distributions of the number of occasions over the F_4 to F_8 generations of the 20 pedigrees (fig. 3) for which there is evidence of segregation at the 5 and 1 per cent levels of significance for flowering time and final height and estimates of k based upon them for P_{Max} and P_{Min}

Character		Flowering time	Final height
Number of significances for ≤0·05	0	2	0
	1	4	4
	2	4	9
	3	4	5
	4	4	0
	5	1	2
	6	1	0
Total of 1 or more		18	20
	k	4–7 (3–3)	∞ (4–5)
Number of significances for ≤0·01	0	6	3
	1	5	5
	2	3	9
	3	3	2
	4	3	1
	5	0	0
Total of 1 or more		14	17
	k	3–3	4–5

ficance level two pairs of estimates are given, the second pair in brackets being those obtained if we omit F_2's for which the evidence for segregation is a single 5 per cent significance. The only effect of any consequence of doing this is to bring the indeterminate estimate for final height into line with the other estimates.

6. CONCLUSIONS

The most significant finding of this study is the unambiguous demonstration that the number of effective factors increases steadily over successive generations of selfing following an initial cross. Although we reached this same conclusion from our earlier study of a completely independent cross (Jinks and Towey, 1976) we did not then have the range of internal checks and controls of both the material and the method that the present data provide. Thus we have been able to demonstrate the soundness of the data and our procedures by using a known single major gene difference controlling anthocyanin production. We have also been able to demonstrate the relatively low estimate of the number of effective factors in the F_2 by three different procedures which use all the data available in one form or another. In so doing we have shown that the low estimate in the F_2 is quite unrelated to the generation used to assess the F_2. Since all the data used to estimate the number of effective factors in the later generations were also used to estimate the number in the F_2 there can be no explanation of the higher estimates in the F_3, F_4, F_5 and F_6 other than that they are genuinely higher.

The explanation of this increase in the number with generation rests upon the nature of effective factors (Mather and Jinks, 1971) and the fact

that strong linkage disequilibrium can be initially generated in a cross between a pair of lines and subsequently resolved during successive rounds of recombination. That there is linkage disequilibrium for the genes controlling the variation in FT and FH in this cross has been demonstrated on a number of occasions (see Perkins and Jinks, 1970). In these circumstances estimates of k, the number of factors based on the early generations of a cross, are expected to be low. In these data, for example, each gene or effective factor detected in the F_2 had become on average ten by the F_6 with no indication that any limit to the increase had been reached.

There are further reasons why the high estimates in the F_6 may still be an underestimate, namely, that the data were not collected from an optimally designed experiment for this purpose or from plants grown under optimal conditions. Thus plots not single plants were the units of randomisation and much of the data was collected during the initial stages of developing a new experimental site. Both reduced the sensitivity with which segregation could be detected and hence our estimates of k.

For the F_2, but not for the other generations, we were able to compare three different ways of using genotype assay for estimating k. Two of these, namely those based upon $p = 2$ and $p = 4$, are strictly comparable in that each estimate is based on the same total number of plants. The third based on lineages, combines all the information from all the generations. The latter method, as might be expected, gave higher estimates than those obtained for $p = 2$, and for FT but not FH the estimates for $p = 4$ were also higher.

The present analyses confirm the overwhelming superiority of genotype assay over all other methods of estimating the number of effective factors that are available for all except a few species. The better of the previous estimates for FT and FH for the V1 × V5 cross have ranged between three and seven and the highest previous estimates based upon F_{10} inbreds were seven for FT and five for FH (Perkins and Jinks, 1973).

7. REFERENCES

BREESE, E. L. 1954. Continuous variation in higher plants. Ph.D. Thesis, University of Birmingham.

BUCIO ALANIS, L., PERKINS, J. M., AND JINKS, J. L. 1969. Environmental and genotype-environmental components of variability. V. Segregating generations. *Heredity, 24,* 115-127.

JINKS, J. L., AND TOWEY, PHILOMENA. 1976. Estimating the number of genes in a polygenic system by genotype assay. *Heredity, 37,* 69-81.

MATHER, K., AND JINKS, J. L. 1971. *Biometrical Genetics.* 2nd Edition. Chapman and Hall, London.

MATHER, K., AND VINES, A. 1952. The inheritance of height and flowering time in a cross of *Nicotiana rustica,* from *Quantitative Inheritance,* 49-80. H.M.S.O., London.

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

PERKINS, J. M., AND JINKS, J. L. 1973. The assessment and specificity of environmental and genotype-environmental components of variability. *Heredity, 30,* 111-126.