PREDICTING THE PROPERTIES OF RECOMBINANT INBRED LINES DERIVED BY SINGLE SEED DESCENT

J. L. JINKS and H. S. POONI

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

SUMMARY

The genetical components of family means and variances, simply and reliably estimated in the first few generations of any breeding programme initiated from a cross between two inbred lines, are all that is required to predict the distribution of the inbred lines that can be derived by single seed descent from the F_2 of such a cross. We can, therefore, determine the probability of obtaining inbreds that fall outside of the parental range or of exceeding the F_1, if it shows heterosis, by any amount we care to specify.

While epistasis, genotype × environmental interactions and linkage may at first sight appear to seriously disturb any such prediction they can all be accommodated. In practice only epistasis is expected and found to produce disturbances of a sufficient magnitude to justify making allowance for its presence in making these predictions. This is because it leads to asymmetry in the distribution of derived inbreds relative to the initial inbred parental means.

The predictive power of this new approach is demonstrated on two complete breeding cycles, each initiated by a cross between two pure-breeding varieties of *Nicotiana rustica* and culminating in the production of over 100 recombinant, pure-breeding lines. The ability of the analysis to discriminate between two crosses on the basis of their probabilities of producing inbred lines that fall outside of their parental range is also demonstrated. By using analyses of this kind there is no reason why we need ever go beyond the F_2 of an inbreeding programme without a fairly clear idea of the outcome.

1. INTRODUCTION

FOR any continuously varying character the expected mean and variance of all possible pure-breeding lines, derived by inbreeding by single seed descent following an initial cross between a pair of such pure-breeding lines, can be specified in terms of the components of means and variances of biometrical genetics (Mather and Jinks, 1971). For example, if an additive (D) and dominance (H) genetic and additive environmental (E) model of gene and environmental action is adequate, the expected mean is m, the mid-parent value, and the expected variance is $D+E$. By adding further components to these simple expectations we can accommodate the effects of non-allelic interaction, genotype × environmental interactions and linkage. The components in these expectations can either be estimated from relatively simple experiments or, if they are too difficult to estimate, the direction and often the likely magnitude of the bias they would introduce can be predicted with sufficient precision for all practical purposes. From the predicted mean and variance we can determine many of the properties of the recombinant inbred lines that can be derived by a selfing programme, based on an initial cross between a pair of pure-breeding lines, at a relatively early stage in the programme. We can also predict the relative probabilities with which

different pairwise crosses will produce inbred lines with particular properties and hence provide an objective criterion for choosing between them.

2. The simple case

In the simple case already referred to, m is defined and estimated as the mean of the original pair of parents (P_1 and P_2) from which the inbreeding programme is initiated, that is

$$m = \tfrac{1}{2}(\bar{P}_1 + \bar{P}_2)$$

and m is also the mean of all possible pure-breeding lines ($\bar{F}_\infty$) that can be derived from the cross between them providing that the initial parents and the derived inbreds are raised in the same environment. If they are raised in different environments these means would in general differ by an amount, e, which would be the difference in the additive environmental values (e_j) of the two environments. However, in any well conducted experiment the two pure-breeding parents, P_1 and P_2, would not only be grown at the outset of the breeding programme when the components were being estimated, but they would also be grown in the final assessment of the derived inbred lines as a control. Since they are pure-breeding lines this presents no problems. The components of the means are then measured relative to the mid-parental value in the environment in which they are estimated and the derived inbreds whose properties we are attempting to predict are measured relative to the mid-parent value in the environment in which they are assessed. Any difference between the additive environmental values of the two environments will then not affect the outcome.

The additive genetic component of the expected variance, D, can be readily estimated from a variety of breeding programmes initiated by a cross between two pure-breeding lines (Mather and Jinks, 1971). The standard P_1, P_2, F_1, F_2 and first backcross, B_1 and B_2 families provide the most rapid estimates (Mather and Vines, 1952) and a triple test cross in which the F_2 is backcrossed to P_1, P_2 and F_1 provides the most reliable estimate (Kearsey and Jinks, 1968; Jinks and Perkins, 1970; Pooni and Jinks, 1976).

The additive environmental component of the expected variance which would be appropriate for predicting the distribution of the means of the derived inbred line is obviously as much a function of the experimental design, the unit of randomisation and the level of replication, as of the environment itself. As pure-breeding lines they can be multiplied to any level required and the within line variation will be entirely environmental in origin. If r units, whether individuals or plots, have been independently randomised for assessing each line, the appropriate E for the distribution of line means will be E_w/r where E_w is the within line, replicate variance averaged over lines.

E_w can either be estimated from the initial pure-breeding parental lines, P_1 and P_2, or E can be reduced to relative unimportance by having a sufficiently large value for r. The latter can be achieved without necessarily increasing the total experimental size by making the unit of randomisation correspondingly small. One can then use the estimate of E_w from the parental lines to determine the experimental design for assessing the derived inbred lines that will most efficiently minimise the value of E.

If, as would usually be the case, the primary interest lay in the repeatable properties of the derived inbreds, that is, those which would be observed on average over many environments, the predicted variance of these inbreds would be D alone without any environmental component.

3. PREDICTING SOME USEFUL PROPERTIES

When the simple model is adequate we can estimate from the first few generations of a breeding programme all the components required to predict the distribution of pure-breeding lines that will emerge from the programme by inbreeding. In general, however, a breeder is more likely to be concerned with the proportion of inbred lines that will prove to be superior to the initial pair of parents, or to their F_1 if it displays heterosis, than with the mean and variance of all possible inbreds. It would be more relevant, therefore, to predict the probability of producing inbreds which fall outside of the parental range or exceed the F_1.

To have a higher score than P_1, the higher scoring parent, or a lower score than P_2, the lower scoring parent, a recombinant inbred must deviate from the mean m by more than plus or minus $[d]$, respectively. The probability of this occurring in a random sample of all possible inbreds from a cross between P_1 and P_2 is given by the probability integral

$$2 \int_{[d]}^{\infty} F(x).dx$$

If the distribution of these inbreds is approximately normal this probability integral will approximate to the two-tail normal probability integral corresponding with the value:

$$\frac{[d]}{\sqrt{D+E}}$$

In most cases, however, we shall omit E because we are interested only in the repeatable performances of the inbreds and the assessments of the recombinant inbreds will be designed to minimise any non-heritable differences between the line means.

Where the F_1 displays heterosis the probability of obtaining recombinant inbreds superior to the F_1 will be

$$\int_{[h]}^{\infty} F(x).dx$$

and again with an approximately normal distribution among the inbreds this will approximate to the one-tail normal probability integral corresponding with the value.

$$\frac{[h]}{\sqrt{D+E}}$$

4. COMPLICATIONS

(i) *Genotype × environmental interaction*

In the presence of genotype × environmental interactions our estimate of $[d]$ will in fact be $[d]+g_d$ and our estimate of D becomes $D+G_D$ where g_d

and G_D are environmentally dependent components. In general there will be two environments or two sets of environments to be considered; the environment or set in which the components are estimated, and the environment or set in which the recombinant inbreds are assessed relative to some control such as the original parents or their F_1. If the common genotypes do not interact across these two environments the simple prediction will hold. Equally, if the components are estimated over one set of environments and the inbreds are assessed over another set, the simple prediction will not be biased if the interactions across the two sets do not differ in magnitude or kind from those occurring between environments within each set. That is, there will be no bias if every environment used can be regarded as being drawn at random from the same population of environments.

Furthermore, even if there is significant genotype × environmental interaction between the two environments or sets of environments the simple prediction may still hold providing that the cause of the interaction is mainly a linear expansion or contraction of the scale. That is, the phenotypic differences between genotypes are magnified in one environment or set relative to the other with little or no changes in ranking. In practice this seems to be the most likely form the interactions would take unless the environments are particularly diverse.

(ii) Epistasis

In the presence of epistasis our estimate of $[d]$ is unchanged but while the mean of the recombinant inbreds is still m the mean of the original pair of parents is now $m + [i]$. This difference in the expected means, unlike that which can arise from environmental differences, cannot be circumvented by modification of the experimental design. The consequences can be readily seen. If $[i]$ is positive the probability of a derived inbred exceeding the larger scoring parent P_1 will be less than the probability of a derived inbred falling short of the smaller scoring parent P_2. The reverse is, of course, true if $[i]$ is negative. There is no longer symmetry in the expectations. However, m and $[i]$ can be readily estimated from parents F_1, F_2 and first backcross families or a triple test cross. Hence, we can correct for the shift in mean between the original parents and the derived inbreds in making our predictions.

The probability of exceeding P_1 is given by the integral

$$\int_{[d]+[i]}^{\infty} F(x).dx$$

and of falling short of P_2, by the integral

$$\int_{-\infty}^{-[d]+[i]} F(x).dx$$

The variance of the recombinant inbreds will no longer equal D but $D + I$. In general we will be unable to estimate precisely $D + I$ from the earlier generations but unavoidably in the presence of epistasis any estimate we attempt to make of D will be biased by the inclusion of epistatic effects although not of the i type. Given an approximately normal distribution, therefore, the probability of obtaining inbreds that score higher than P_1 will be the one-tail normal probability integral corresponding with the value:

$$\frac{[d]+[i]}{\sqrt{D+I}}.$$

Similarly, for inbreds scoring lower than P_2 it will be the one-tail probability integral corresponding with the value:

$$\frac{-[d]+[i]}{\sqrt{D+I}}.$$

Which one of the two probabilities is the larger will depend on the sign of $[i]$ itself. The sum of the two probabilities will give us the probability of inbreds falling outside of the parental range. This probability will always be underestimated if we fail to make allowance for the presence of a significant $[i]$ component. More important, however, we would fail to predict the inequality between the proportion that would exceed P_1 and that which would fall short of P_2.

(iii) *Linkage*

In the absence of other complications such as epistasis, linkage has no effect on the expected mean of the original parents or of the derived inbreds. It affects, however, the expectation of the additive component of variation D and it does so to an extent that differs between the early generations of a cross and the later generations of derived inbred lines. In the latter D becomes:

$$D_{F\infty} = D\frac{+C}{-R}\Sigma\frac{2(1-2p)}{1+2p}\,d_id_j.$$

In contrast the D of the F_2 from which the inbreeding programme is initiated is:

$$D_{F2} = D\frac{+C}{-R}\Sigma\,2(1-2p)\,d_id_j.$$

Because of the additional opportunities for recombination to reduce the linkage disequilibrium the linkage bias on D is smaller in the advanced than in the early generations. The correct D for predicting the properties of the derived inbreds will therefore be greater than that estimated from the early generations if the linkages are predominantly in the repulsion phase and smaller if they are predominantly in the coupling phase. If linkage is present and undetected the D we would be estimating from the F_2 and back-crosses or from a triple test cross would underestimate the proportion of derived inbred lines which would lie outside of the range of the original parents if repulsion linkages predominate, and overestimate this proportion if coupling linkages predominate. Since we are primarily considering the probability of producing inbreds more extreme than the initial inbred parents, the linkages that will concern us will be mainly in the repulsion phase and hence in general we will be underestimating the probability of success.

The best opportunity for detecting linkage and classifying the predominant phase is provided by the comparison of F_2 and $F_2 \times F_1$ variances as part of the triple test cross analysis (Jinks and Perkins, 1970). On the basis of

this analysis the direction of the bias and to some extent the likely magnitude of the bias in predictions based on the estimate of D can be determined.

5. EXAMPLES

(i) *Cross 1 × 5, 1946-61*

The cross between pure-breeding varieties 1 and 5 of *Nicotiana rustica* was initiated by Mather and Vines at Merton, London (see Mather and Vines, 1952) and from this cross 20 pure-breeding lines were ultimately derived and assessed at Winterbourne, Birmingham (see Perkins and Jinks, 1968; and Jinks, Perkins and Breese, 1969 for details). Each pure-breeding line was independently derived by selfing from a single, randomly chosen F_2 individual. Estimates of D and $[d]$ from the early generations of this cross are available from the 1946 season onwards (Mather and Vines, 1952) but because of a major shift in phenotype on transferring the material from Merton to Winterbourne only the estimates from the 1950 season onwards are useful for predicting the properties of the derived inbred lines as observed at Winterbourne.

In table 1 are listed the estimates of D and $[d]$ for final height (in inches) for the 1950, 51 and 52 seasons which are based essentially upon P_1, P_2, F_1, F_2, B_1 and B_2 generations but supplemented by progressively later generations derived by selfing (Breese, 1954) and for the 1954 season based upon an F_2 triple test cross (Jinks and Perkins, 1970). The two-tail normal probability integral corresponding with the values of $[d]/\sqrt{D}$ give consistently high probabilities and they lead on average to a probability in excess of 80 per cent of deriving inbred lines which lie outside of the parental range.

TABLE 1

Estimates of D, [d] *and* P, *the two-tailed normal probability integral derived from them, for final height. The 1950, 51 and 52 estimates are based on* F_2 *and first backcross families and the 1954 estimates on an* F_2 *triple test cross*

	Estimates		
Seasons	D	$[d]$	P
1950	41·86	2·09	0·75
1951	47·03	1·35	0·84
1952	36·58	0·50	0·93
Mean	41·82	1·31	0·84
1954	32·39	1·40	0·81

By 1953, the 20 independently derived inbred families had reached the F_9 generation having been initiated from the original F_2 in 1946. During the 1953, 54, 59 and 61 seasons these inbred families were assessed alongside the original parental varieties 1 and 5 at the F_9, F_{10}, F_{13} and F_{17} generations respectively. The results of these assessments are summarised in table 2. For final height the four seasons ranked 3rd, 8th, 1st and 14th out of the 16 seasons between 1946 and 1964 that were reviewed by Bucio-Alanis (1966) Bucio-Alanis and Hill (1966) and Bucio-Alanis, Perkins and Jinks (1969). One season, 1954, is therefore average, while the other three are relatively extreme for the development of final height in the 1 × 5 cross. This is reflected in the estimates of m from the mean of the two original

parents and from the mean of the 20 inbred lines in each of the four seasons. Although there are clearly large differences between seasons there is close agreement between these two independent estimates within each season. The simple model is, therefore, adequate and we expect equality in the proportion of inbreds that fall below and above the parental range.

The number of inbred lines which are smaller than the smaller parent (P_2) and larger than the larger parent (P_1) have been separately listed within each season as well as for the data averaged over all four seasons. The proportion of inbred lines that fall outside of the parental range in no

TABLE 2

The observed number of inbred lines that are taller than P_1 and shorter than P_2 and the proportion of lines that fall outside the parental range in assessments made in 1953, 54, 57 and 1961. The same observations on the average performances over all four assessments are also included. The relative values and rankings of the four seasons used for assessments based on the mid-parent and inbred line mean are shown

| Seasons | Observations | | | Inbreds shorter than P_2 | Inbreds taller than P_1 | Proportion outside P_1, P_2 range |
	Mid-parent value	Mean of inbreds	Ranking 1946-64			
1953	39·34	39·70	3	11	5	0·80
1954	43·29	44·56	8	8	10	0·90
1957	34·35	31·38	1	11	5	0·80
1961	52·50	52·25	14	6	7	0·65
Combined	42·37	41·97	10	9	6	0·75

case differs significantly from the prediction. Furthermore, in no case is there a significant deviation from the expectation that the frequency of inbreds which are smaller than P_2 should equal the frequency of inbreds which are larger than P_1, although overall there is an excess of the former. There is, therefore, good agreement between observation and prediction in spite of three of the four environments used for testing the predictions being among the more extreme experienced among the sixteen environments which have been sampled.

(ii) *Cross 1 × 5, 1965-73*

To extend their earlier studies Jinks and Perkins in 1965 initiated a new breeding programme with varieties 1 and 5 on the new, main campus, experimental field. The new site provided a very different environment and it was necessary to re-estimate all the components of the means and variances using the conventional breeding programmes and the triple test cross. The main objective, however, was to obtain a large random sample of true-breeding lines from this cross as quickly and as economically as possible. Inbreeding by selfing was therefore initiated from a random sample of 100 F_2 plants at the rate of two generations per year. In each generation each inbred line consisted of a single plant raised from a randomly chosen single seed produced by selfing the sole representative of the line in the previous generation. Generations were alternated between the glasshouse (October to March) and the experimental field (May to October). No measurements were taken during the inbreeding programme and all assessments of the inbred lines were deferred until they were judged to be more or less true-breeding.

The first assessment was made at F_8 in the 1970 season. Further assessments were made at the F_9 (1971) and F_{11} (1973). Although the inbred lines were assessed for many characteristics the two for which we have complete records over all experiments since 1965 are flowering time and final height.

Estimates of the relevant components of means and variances were made by Jinks and Perkins from triple test crosses and conventional breeding programmes raised in 1965. These estimates are summarised in table 3. The different estimates of D lead to remarkably consistent estimates of P, the probability of obtaining pure-breeding lines that are outside the parental range, for flowering time and to reasonably consistent values for final height. A number of other properties of the genetical variation for flowering time and final height in the new environment emerge from these analyses which are also relevant to our predictions.

TABLE 3

Estimates of D, [d] *and* P *for final height and flowering time based on breeding programmes in 1965*

Source	Final height		Flowering time
		D	
1. F_2 and backcrosses	51·84		31·58
2. TTC on F_2	30·69		25·95
3. TTC on backcrosses	53·55		26·87
Mean	45·36		28·13
		[d]	
1.	5·18		1·31
		P	
1.	0·47		0·82
2.	0·35		0·80
3.	0·48		0·80
Mean	0·43		0·80

The most important of these properties is a small but significant directional epistatic component for both characters (Jinks and Perkins, 1969, 1970; Perkins and Jinks, 1970). Thus [i] which measures the displacement between the mid-parent value and the mean of all possible derived inbred lines (see Section 4 (ii)) is negative and equal to $-1\cdot38$ days for flowering time and $-6\cdot02$ inches (or $-15\cdot29$ cm) for final height in 1965. We expect, therefore, that the mid-parent value which is $m+[i]$ in the presence of such epistasis will underestimate the mean, m, of the derived inbred lines. As a consequence, of the derived inbred lines that fall outside of the parental range, more should exceed P_1 than should fall short of P_2 for these two characters.

There is also evidence of linkage (Jinks and Perkins, 1969; Perkins and Jinks, 1970) and while again the effect is not large it is, as might be expected, in the repulsion phase. Two consequences may be expected (Section 4 (iii)); we shall underestimate the true value of D, (Σd^2), but equally the probability of extreme phenotypes appearing among the recombinant inbreds will be reduced. Our predictions of the probabilities of obtaining inbred lines that fall outside the parental range and the proportion of inbred lines that actually meet this criterion will both be reduced although not necessarily to the same extent. All three estimates of D for each character in table 3 are rank 1 estimates and therefore, have the same linkage bias (Mather and Jinks, 1971).

For repulsion linkages

$$D = \Sigma d^2 - \Sigma 2(1-2p)d_i d_j$$

while the genetical variation among the recombinant inbred lines is

$$\Sigma d^2 - \Sigma \frac{2(1-2p)}{1+2p} d_i d_j$$

Both will be less than the true value of $D = \Sigma d^2$ but the D we estimate will be smaller than the additive genetical variation among the derived inbreds. We shall, therefore, underestimate the probability of obtaining inbred lines which fall outside of the parental range. But this under-estimation will rarely be serious because it is impossible to have more than two genes linked in repulsion without some of them being in coupling and their opposing biases will then be cancelled out (Mather and Jinks, 1971).

TABLE 4

The observed number of inbred lines that are taller or flower later than P_1 and shorter or flower earlier than P_2 and the proportion of inbred lines that fall outside of the parental range for final height and flowering time. The observations are based on assessments made in 1970, 1971 and 1973. The 1971 observations are combined over eight contrasting environments

| | Final height | | Proportion |
| | Inbreds shorter | inbreds taller | outside P_1, |
Season	than P_2	than P_1	P_2 range
1970	16	45	0·74
1971 (8 environments)	17	38	0·67
1973*	19	31	0·63
Mean	17	38	0·67

| | Flowering time | | |
| | Inbreds earlier | Inbreds later | |
	than P_2	than P_1	
1970	15	57	0·87
1971 (8 environments)	11	55	0·80
1973*	5	47	0·65
Mean	10	53	0·77

* Based upon 80 derived inbred lines.

The results of the selfing programme by single seed descent are summarised in table 4. Of the initial 100 F_2 plants chosen to initiate the inbreeding programme 82 were still represented by an inbred line when assessments commenced at the F_8 in 1970 and no further losses had occurred when the current assessments had been completed at the F_{11} in 1973. Details of the experimental design used in these assessments are given by Eaves and Brumpton (1972) and Perkins and Jinks (1973). The number of inbred lines that were shorter or flowered earlier than variety 1 (P_2) or were taller and flowered later than variety 5 (P_1) in each of these assessments are tabulated in table 4. The proportion of the 82 inbreds that fell outside of the parental range for each character are also shown. These observations differ to some extent over the three seasons but they are consistently of the same order of magnitude and the relative positions of the two characters remain the same. The 1971 season provides results of the greatest reliability and

generality because they are based on the combined assessments in eight different environments (Perkins and Jinks, 1973).

For flowering time there is remarkably good agreement between the observations and the predictions. As predicted, about 80 per cent of the derived inbreds fall outside of the parental range and as expected in the presence of epistasis considerably more inbreds flower later than P_1 than flower earlier than P_2 and the mean flowering time of the derived inbreds is later than the mid-parent value in every one of the 10 environments.

TABLE 5

Amended probabilities allowing for the level of epistasis observed in 1965. The observed proportions are given for comparison

Predicted proportions Source	Final height		Proportion outside P_1, P_2 range
	Proportion shorter than P_2	Proportion taller than P_1	
1	0·06	0·55	0·61
2	0·02	0·56	0·58
3	0·06	0·55	0·61
Observed proportions Season			
1970	0·20	0·55	0·74
1971	0·21	0·46	0·67
1973	0·24	0·39	0·63

Predicted proportions Source	Flowering time		Proportion outside P_1, P_2 range
	Proportion earlier than P_2	Proportion later than P_1	
1	0·32	0·51	0·82
2	0·30	0·51	0·81
3	0·30	0·51	0·81
Observed proportions Season			
1970	0·18	0·70	0·87
1971	0·13	0·67	0·80
1973	0·06	0·59	0·65

For final height the agreement with the predictions is not quite as good. Rather more derived inbreds fall outside of the parental range than we predict, but this is expected in the presence of epistasis as is also the greater number of inbreds that are taller than P_1 than are shorter than P_2. Also expected for the same reason is the greater mean final height of all the inbreds relative to the mid-parent value, which we observed in nine of the 10 environments.

Epistasis is clearly having a marked effect and we must amend our predictions using the estimates of $[i]$ from 1965 (Section 4 (ii)). The amended probabilities are given in table 5 and the observed proportions are included for comparison. The inclusion of $[i]$ has improved the agreement between prediction and observation in every case but most markedly for the overall proportion of inbreds which are outside of the parental range for

final height. However, although we have successfully predicted the direction and order of magnitude of the asymmetry it is more extreme than predicted for flowering time and less extreme than predicted for final height. The reason is simply that the estimate of $[i]$ made in 1965 was below average for flowering time and above average for final height over the period of assessment and the predicted asymmetry is very sensitive to the magnitude of $[i]$. Previous investigations of the 1×5 cross and other crosses between pure-breeding lines of *Nicotiana rustica* suggest that the magnitude of the $[i]$ type epistasis changes with the environment, achieving its greatest value in extreme environments.

Since the overall proportions of inbreds which fall outside of the parental range agree well with predictions, the low level of repulsion linkage we detected is not having any noticeable effects and, of course, linkage itself cannot contribute to the asymmetry.

(iii) *Other examples*

Although the two sets of data from the 1×5 cross illustrate our approach at two different levels of complexity the probability of obtaining inbreds which lie outside of the parental range is in all cases high. We will, therefore, consider a case where this is not so. Such a case is provided by the cross between inbred lines B2 and B35 (Jayasekara and Jinks, 1976). These two inbred lines were chosen from among the 82 inbreds derived from the 1×5 cross because they were among the high and low extremes, respectively, for final height. But apart from differences in distribution of the alleles for which varieties 1 and 5 differ, the $B2 \times B35$ and the 1×5 cross segregate for the same allelic differences at the same loci. In particular their respective F_1's and F_2's are expected to be, and are observed to be, indistinguishable in mean and variance.

From the $B2 \times B35$ crossing programme we have an estimate of $[d]$ and two estimates of D for each character, from which we can predict the proportion of inbreds which are expected to fall outside of the B2, B35 range. These are summarised in table 6 for the two characters final height and flowering time. Since B2 and B35 were originally chosen for their extreme heights it is not surprising that we predict a probability of only 0·001 to 0·05 of obtaining more extreme inbreds for this character. It seems likely that the first of these probabilities is an underestimate because it is based on an atypically low estimate of D. The probability of obtaining more extreme inbreds for flowering time also take the relatively low values of 0·14 and 0·16.

We have not yet had time to extract new inbreds from the $B2 \times B35$ cross. But in theory and in practice as far as it has gone, the inbreds that emerge should have the same mean and variance as those from the 1×5 cross. We can therefore, compare our predictions with the inbreds derived from the 1×5 cross using the three assessments referred to in the last section. The proportion of inbred lines which we observe to be more extreme than B2 and B35 are in remarkably good agreement with our predictions for both characters in each of the three assessments (table 6). For final height the frequency of inbreds smaller than B35 is equal to the frequency of inbreds taller than B2 and this coincides with our failure to detect significant epistasis in these data (Jayasekara and Jinks, 1976). In contrast there is marked asymmetry for flowering time, which we expect since there is significant epistasis for this character in these data.

6. Conclusions

In this paper we have set out to show that the kind of genetical components that can be most readily and reliably estimated in the early stages of any breeding programme initiated from a cross between two inbred lines can be used to predict the distribution of inbred lines that can be derived by single seed descent from individuals of the F_2 of such a cross. With this information we can determine the probability of obtaining inbreds that fall outside the initial parental range or for a character showing heterosis the probability of exceeding the F_1. Equally, of course, by the same procedures we can predict the probability of obtaining inbreds that fall outside of the parental range or exceed the F_1 by any amount we care to specify.

TABLE 6

Estimates of D, [d] *and* P *for the* B2 × B35 *($P_1 × P_2$) cross and the observed frequency of inbreds taller or later flowering than B2 and shorter or earlier flowering than B35*

	Estimates	Final height	Flowering time
	D	9·98	21·00
		29·02	23·48
	[d]	10·35	6·81
	P	0·001	0·14
		0·05	0·16

Observations	Inbreds smaller than P_2	Inbreds larger than P_1	Proportion outside P_1, P_2 range	Inbreds earlier than P_2	Inbreds later than P_1	Proportion outside P_1, P_2 range
1970	0	0	0·00	2	5	0·09
1971	0	0	0·00	3	16	0·23
1973	3	3	0·07	3	13	0·20

While epistasis, genotype × environmental interactions and linkage may at first sight appear to seriously complicate the simple prediction, they can be accommodated. The epistasis component [i] estimated from F_2 and backcross families or from an F_2 triple test cross is all that is required to correct for epistatic effects, although greater accuracy might be achieved if the magnitude and direction of the epistatic bias in the estimate of D were known. Rather than detecting and allowing for genotype × environmental interactions their disturbance of the predictions can be minimised. This can be achieved by making all predictions on the basis of the relative performances of the derived inbreds and their initial parents and F_1 when grown in the same environment. Linkage can be readily detected and its predominant phase determined as part of an F_2 triple test cross programme. This will alert us to the direction of any bias but its effect is minimal because it biases the estimates of our predictions in the same direction, although not to the same extent, as it biases the actual distribution of recombinant inbreds. In practice, only epistasis has so far produced biases of sufficient magnitude to justify correcting the predictions for its presence and even then it is because of the asymmetry it leads to in the proportions of upper and lower extreme inbreds rather than in the overall proportion of inbreds that fall outside of the parental range.

The predictive power of this approach has been tested on *Nicotiana rustica*

breeding programmes that extend over 25 years and three locations. The programmes include two complete cycles of crossing followed by selfing, such that each F_2 individual is represented at the end of each cycle by a single pure-breeding line. While the sophistication of the analysis and predictions differ between these two cycles they leave little doubt that valuable predictions can be made about the inbred lines that can be derived by single seed descent from the F_2 of a cross between two inbred lines from the early generations of the breeding programme. And we have shown by the comparative analysis of two contrasting crosses, 1 × 5 and B2 × B35 that it is possible to distinguish between crosses with high and low probabilities of producing inbred derivatives which exceed their parents in performance in the upper or lower direction.

If we combine the procedures described in this paper with those for predicting the upper and lower extreme inbreds that might be expected from an inbreeding programme, which can also be based on information from an F_2 triple test cross (Jinks and Perkins, 1972), there is no reason why we need ever go beyond the F_2 of an inbreeding programme without a fairly clear idea of the final outcome.

7. REFERENCES

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

BUCIO-ALANIS, L. 1966. Environmental and genotype-environmental components of variability. I. Inbred lines. *Heredity, 21*, 387-397.

BUCIO-ALANIS, L., AND HILL, J. 1966. Environmental and genotype-environmental components of variability. II. Heterozygotes. *Heredity, 21*, 399-405.

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

EAVES, L. J., AND BRUMPTON, R. J. 1972. Factors of covariation in *Nicotiana rustica. Heredity, 29*, 151-175.

JAYASEKARA, N. E. M., AND JINKS, J. L. 1976. Effect of gene dispersion on estimates of components of generation means and variances. *Heredity, 36*, 31-40.

JINKS, J. L., AND PERKINS, JEAN M. 1969. The detection of linked epistatic genes for a metrical trait. *Heredity, 24*, 465-475.

JINKS, J. L., AND PERKINS, JEAN M. 1970. 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. 1972. Predicting the range of inbred lines. *Heredity, 28*, 399-403.

JINKS, J. L., PERKINS, JEAN M., AND BREESE, E. L. 1969. A general method of detecting additive, dominance and epistatic variations for metrical traits. II. Application to inbred lines. *Heredity, 24*, 115-127.

KEARSEY, M. J., AND JINKS, J. L. 1968. A general method of detecting additive, dominance and epistatic variation for metrical traits. I. Theory. *Heredity, 23*, 403-409.

MATHER, K., AND JINKS, J. L. 1971. *Biometrical Genetics.* 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*, ed. E. C. Reeve and C. H. Waddington, pp. 49-80. H.M.S.O., London.

PERKINS, JEAN M., AND JINKS, J. L. 1968. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity, 23*, 339-356.

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

POONI, H. S., AND JINKS, J. L.   1976.   The efficiency and optimal size of triple test cross designs for detecting epistatic variation, *Heredity*, *36*, 215-227.