

The use of doubled haploids in barley breeding.

3. An assessment of multivariate cross prediction methods

W. Powell*, P. D. S. Caligari*,
J. W. McNicol* and J. L. Jinks†

* Scottish Crop Research Institute, Pentlandsfield,
Roslin, Midlothian, EH25 9RF, U.K.

† Department of Genetics, University of Birmingham,
P.O. Box 363, Birmingham, B15 2TT, U.K.

Following a previous report on the use of the univariate method of predicting the properties of inbred lines produced by single seed descent, predictions are now reported for two or more characters simultaneously. Bivariate, trivariate and quadrivariate predictions have been made using genetical parameters estimated from F_3 and doubled haploid (DH) populations. The additive genetic and phenotypic correlations between pairs of characters were compared for their efficiency in generating predictions. The phenotypic correlations did not appear to result in less accurate predictions.

In general, estimates of genetical parameters obtained from F_3 families resulted in a more accurate ranking of crosses than the corresponding DH system. An alternative method of ranking crosses based on the proportion of lines falling into pre-defined phenotypic classes was examined. Although preliminary, this form of “progeny testing” based on F_3 families holds promise as a simple method of identifying crosses with the most potential.

INTRODUCTION

In a previous paper (Caligari, Powell and Jinks, 1985) it has been shown that genetical parameters obtained from F_3 or doubled (DH) families could successfully be used for univariate cross prediction purposes. Although it is important to examine the results for each character independently, breeding programmes are rarely based on single characters. The proportion of inbred lines satisfying two or more criteria simultaneously cannot be calculated simply as the product of two or more probabilities obtained independently, unless there is no genetic correlation between them. However, Pooni and Jinks (1978) have described procedures whereby joint predictions for two or more characters can be obtained. They showed that only one further parameter was necessary, namely the additive genetic correlation between each pair of characters. Their method is based on the numerical technique of Owen (1956) for estimating bivariate and trivariate probabilities. In this paper genetical parameters estimated from F_3 and DH samples are

used to calculate bivariate and multivariate probabilities, and the expected frequencies are compared with those observed in single seed descent (SSD) populations.

MATERIALS AND METHODS

A detailed account of the experiment, from which are derived the results reported, is given in Caligari, Powell and Jinks (1985) so only a brief account is given here. The following 5 pairs of crosses between barley cultivars were studied:

Golden Promise × Mazurka	(TT1)
Universe × Mazurka	(TT2)
Golden Promise × Ark Royal	(TT3)
BH4/143/2 × Ark Royal	(TT4)
Clipper × Ymer	(TT5).

Twenty DH lines were obtained from F_1 hybrids of the five spring barley crosses. In addition 20 F_3 families were obtained together with a sample of 40 single seed descent lines (SSD) from each cross. The material was grown in a

randomised, replicated experiment in 1983 at the Murrays Farm, East Lothian.

To calculate bivariate probabilities five genetical parameters need to be estimated: the mean of character 1 (m_1), the mean of character 2 (m_2), the corresponding additive genetic variances (D_1 and D_2) and the additive genetic correlation ($D_{1,2}$). The means of the F_3 and DH generations were used to provide estimates of m , the mean of all possible homozygotes. In the case of F_3 samples it is necessary to assume that non-additive effects i.e., $\frac{1}{4}[h]$ and $\frac{1}{16}[I]$ are negligible. The additive genetical variance D or $(D+I)$ for the DH samples was calculated from the between lines mean square using the appropriate expected mean squares for the experiment. Estimates of D may also be obtained from the F_3 generation and in this case it must be assumed that $\frac{1}{8}H$ is negligible. For trivariate predictions it is necessary to estimate three means (m_1, m_2, m_3) and three variances (D_1, D_2, D_3) together with the three pairwise correlations ($D_{1,2}, D_{1,3}$ and $D_{2,3}$) between the characters. For more than three characters the number of parameters which need to be estimated increases correspondingly. Thus for n characters, n means, n variances and $n(n-1)/2$ pairwise correlations must be estimated.

The additive genetic correlations were calculated by performing a one way analysis of covariance for each pair of characters. Such an analysis provides the between families and within families covariance. Estimates of the additive genetic component common to a pair of characters can then be obtained from the expected mean products. The additive genetic covariance thus derived can subsequently be used to obtain the additive genetic correlation, e.g. the additive genetic correlation between characters height (Ht) and thousand grain weight (TGW) is obtained as:

$$\text{Additive genetic correlation} = \frac{\text{Additive genetic covariance (Ht/TGW)}}{\sqrt{D_{\text{Ht}} \cdot D_{\text{TGW}}}}$$

The predictions were calculated using a program developed at SCRI which is based on an algorithm by Schervish (1984) to derive multivariate normal probabilities. This algorithm incorporates Owen's method but is not restricted to two variables. The SCRI program, therefore, not only removes the effort of tedious calculations and reference to tables but removes the constraint implied by Pooni and Jinks (1978) that cross predictions are practical for the bivariate and trivariate cases only. In this sense the SCRI program provides a more powerful tool than has hitherto been available to plant breeders for cross prediction work. Various combinations of characters will be considered but these will be based on the same nine characters as described by Caligari *et al.* (1985).

The observed numbers, out of 40, in the SSD populations and predicted numbers of lines from the F_3 and DH populations falling into four of the possible phenotypic classes are given in table 1 for the TT3 cross. The 4 phenotypic classes considered are:

$>P_1 \quad >P_1 \quad >P_1 \quad <P_2 \quad <P_2 \quad >P_1$
Char 1 Char 2 Char 1 Char 2 Char 1 Char 2
and

$<P_2 \quad <P_2$
Char 1 Char 2

Various combinations of characters are considered and the scores for P_1 (the highest scoring parent) and P_2 (the lowest scoring parent) are given below:

	P_1	P_2
Height in cm (Ht)	66.58	45.70
Thousand grain weight in g (TGW)	53.88	47.80
Main Stem weight in g 10^{-2} (MSW)	161.79	144.39
Maturity (Mat)	7.30	4.40
Awn emergence (AE)	15.90	10.90
Single Plant yield in g (SPY)	13.06	8.61

Table 1 The observed numbers and those predicted from F_3 and DH samples, using the bivariate method, to fall into four phenotypic categories for the cross Golden Promise \times Ark Royal (TT3)

	$>P_1$ Obs	$>P_1$ F_3	$>P_1$ DH	$>P_1$ Obs	$<P_2$ F_3	$<P_2$ DH	$<P_2$ Obs	$>P_1$ F_3	$>P_1$ DH	$<P_2$ Obs	$<P_2$ F_3	$<P_2$ DH
Ht/TGW	2	1	2	0	0	1	0	0	1	5	1	2
Ht/MSW	3	7	4	0	1	0	0	0	0	5	0	2
Ht/Mat	2	0	2	1	0	0	0	0	0	2	0	2
Ht/AE	0	1	2	0	0	0	1	1	0	2	0	0
Ht/SPY	1	1	2	0	0	0	0	0	0	3	2	1
SPY/AE	0	2	0	0	0	3	4	3	6	1	0	0
SPY/Mat	1	1	1	2	1	10	1	2	4	3	2	12

Table 2 The predicted numbers based on F_3 and DH samples together with the observed number of lines for the trivariate distribution of Height (Ht), Thousand grain weight (TGW) and Single plant yield (SPY) in the Golden Promise \times Ark Royal cross (TT3)

Ht	TGW	SPY	DH		F_3		SSD
			Prob.	Exp.	Prob.	Exp.	Obs
>60	>50	>12	0.1248	5	0.0924	4	3
>60	>50	<8	0.0069	0	0.0171	1	1
>60	<40	>12	0.0100	0	0.0000	0	0
>60	<40	<8	0.0101	0	0.0000	0	0
<48	>50	>12	0.0167	1	0.0261	1	0
<48	>50	<8	0.0028	0	0.0169	1	1
<48	<40	>12	0.0054	0	0.0000	0	0
<48	<40	<8	0.0148	1	0.0000	0	2

In general, the observed and predicted numbers from both the F_3 and DH samples agree reasonably well. The bivariate prediction method can be extended to predict the properties of inbreds in respect of three and four characters. Trivariate predictions have been made for various combinations of characters, those for height, thousand grain weight and single plant yield for the TT3 cross are given in table 2. Also included in table 2 are the standards imposed for each of the eight phenotypic categories. The sample size of 40 SSD lines which provide the observed frequencies is clearly too small to permit reliable statistical tests but the trivariate predictions appear to be of the correct relative magnitude. Quadrivariate predictions have been made for several sets of four variates and that for the combination of height, thousand grain weight, single plant yield and awn emergence, based on genetical parameters, obtained from DH in the TT5 cross, is shown in table 3. The predicted quadrivariate proportions of inbreds with scores falling outside the limits set have been obtained for the 16 combinations of interest together with the observed numbers. In this case some of the phenotypic categories agree reasonably well, but for others *e.g.*, Ht > 60; TGW > 50; SPY < 8; AE > 10 the estimated probability of 0.00708 is clearly disparate from the observed frequency of 4 out of 40 *i.e.*, 0.10. This pattern of some departures between observed and predicted frequencies was found in all the combinations studied.

An interesting point to emerge during the computation of higher order predictions related to the role of genetic correlations in the prediction equation. The method of calculating genetic correlations based on components of variance and covariance does not restrict the observed distribution to the range -1 to $+1$ in the normal way. Therefore, correlations greater than one or less

Table 3 The predicted numbers based on DH samples together with the observed number of lines in the Clipper \times Ymer (TT5) cross for the quadrivariate distribution of Ht, TGW, SPY and AE

Ht	TGW	SPY	AE	Prob	Exp	Obsv
>60	>50	>12.0	>10	0.09134	4	4
>60	>50	>12.0	<4	0.00366	0	0
>60	>54	<8	>10	0.00708	0	4
>60	>50	<8	<4	0.00069	0	0
>60	<40	>12	>10	0.00000	0	0
>60	<40	>12	<4	0.00000	0	0
>60	<40	<8	>10	0.00000	0	0
>60	<40	<8	<4	0.00000	0	0
<48	>50	>12	>10	0.00013	0	0
<48	>50	>12	<4	0.00064	0	0
<48	>50	<8	>10	0.00389	0	1
<48	>50	<8	<4	0.03778	2	0
<48	<40	>12	>10	0.00000	0	0
<48	<40	>12	<4	0.00000	0	0
<48	<40	<8	>10	0.00000	0	1
<48	<40	<8	<4	0.00000	0	1

than minus one are frequently obtained and are then assumed to be 1 or -1 . When three or more variates are considered such approximations do not represent a true picture of the relationships between the characters and hence do not permit the calculation of multivariate probabilities. Phenotypic correlations were therefore used to calculate trivariate and quadrivariate probabilities. Although phenotypic correlations contain environmental as well as genetic components the predictions obtained are, in fact, reasonably accurate. In the case of the bivariate predictions the correlation coefficients between observed and expected numbers based on additive genetic and phenotypic correlations are given below:

	DH	F_3
From the genetic correlation	0.44	0.37
From the phenotypic correlation	0.38	0.51

As can be seen, an improvement in the correlation between observed and expected is produced when the phenotypic correlations replaces the genetic correlation in the F_3 samples. On the other hand, a lower correlation coefficient is produced in the DH sample when the phenotypic correlation is used in the bivariate prediction equation, though it is not greatly so.

Although the multivariate predictions within a cross are of interest, the concept of ranking crosses is of greater importance to plant breeders. The ranking was examined further with reference to bivariate predictions and is illustrated by considering height (Ht) with six other characters:

height:thousand grain weight	(Ht/TGW)
height:main stem weight	(Ht/MSW)
height:maturity	(Ht/Mat)
height:awn emergence	(Ht/AE)
height:grain number	(Ht/GN)
height:single plant yield	(Ht/SPY)

By setting standards for the seven characters of interest it is then possible to make comparisons between the five crosses. The five genetical parameters necessary for bivariate predictions together with the seven standards were used to make predictions for the six combinations of characters. Such predictions based on F_3 samples and DH samples for the five crosses in each of the four phenotypic classes, were ranked within each phenotypic class. The predicted ranking for each character combination was then compared with the observed ranking in the SSD populations. For this purpose Spearman's rank correlation technique was used. The 24 correlation coefficients obtained, based on F_3 data, are given in table 4. In general the correspondence between observed and expected ranking is reasonable but poor correlations were observed in at least four of the cases.

Table 4 Rank correlation coefficients between the predicted and observed cross rankings, based on genetical parameters obtained from the F_3 sample

	$>P_1 > P_1$	$>P_1 < P_2$	$<P_2 > P_1$	$<P_2 < P_2$
Ht/TGW	-0.188	-0.150	0.525	0.100
Ht/MSW	-0.025	0.775	0.375	0.575
Ht/Mat	0.800	0.775	0.850	0.600
Ht/AE	0.875	1.000	0.825	0.850
Ht/GN	0.175	0.725	0.225	0.200
Ht/SPY	0.000	0.525	0.375	0.125

The rank correlation coefficients between the observed and predicted rankings based on DH samples are given in table 5. Low correlation

Table 5 Rank correlation coefficients between the predicted and observed cross rankings, based on genetical parameters obtained from the DH sample

	$>P_1 > P_1$	$>P_1 < P_2$	$<P_2 > P_1$	$<P_2 < P_2$
Ht/TGW	-0.600	-0.075	0.750	0.263
Ht/MSW	0.125	0.850	0.750	0.375
Ht/Mat	0.300	0.825	0.750	0.225
Ht/AE	0.375	1.000	0.200	0.850
Ht/GN	-0.600	0.300	0.600	0.750
Ht/SPY	-0.150	0.275	0.125	0.375

coefficients are again observed for combinations involving TGW, SPY and MSW. Furthermore the correlations are, in general, lower than the corresponding estimates based on the F_3 sample. It would therefore appear that on average the F_3 sample provides a more accurate method of ranking crosses based on bivariate predictions.

The bivariate prediction procedure involves calculating the means and variances for each pair of characters and also the genetic correlation between these characters. It is therefore of practical and theoretical significance to compare the relative importance of the means, variances and correlations in the bivariate prediction method. For this purpose multiple regression procedures were employed (Caligari *et al.*, 1985). The predicted rankings of the crosses were regressed against the means (m_1 and m_2), variances (D_1 and D_2) and the additive genetic correlation ($D_{1,2}$) calculated from the SSD populations. The coefficients of determination (R^2) calculated from the regression analyses are given in table 6. With certain exceptions the R^2 's are relatively low and the predicted rankings would appear to depend on all three genetic components *i.e.*, means, variances and correlations.

The bivariate prediction methods outlined in this paper involved the calculation of five genetic parameters and all five parameters appear to be necessary to obtain accurate predictions. Therefore the method involves a number of computational tasks which in some cases *e.g.*, Ht/TGW, do not result in very accurate rankings of crosses. Under such circumstances perhaps a compromise to achieve simplicity may be justified. The simplest approach to prediction is to calculate the observed proportion of lines falling into each of the four bivariate phenotypic classes in the F_3 and DH samples. It is then possible to rank the crosses based simply on the proportions observed in both the samples. Rank correlation coefficients can subsequently be used to compare the predicted and observed rankings. The correlation coefficients

Table 6 The coefficients of determination (R^2) for m_1m_2 , D_1 , D_2 and $D_{1,2}$ in the four phenotypic classes

	$>P_1 > P_1$			$>P_1 < P_2$			$<P_2 > P_1$			$<P_2 < P_2$		
	m_1m_2	D_1D_2	$D_{1,2}$	m_1m_2	D_1D_2	$D_{1,2}$	m_1m_2	D_1D_2	$D_{1,2}$	m_1m_2	D_1D_2	$D_{1,2}$
Ht/TGW	35.13	31.33	17.41	33.23	64.45	3.34	—	—	—	93.91	37.28	78.62
Ht/AE	21.88	71.95	32.43	97.51	97.51	14.51	61.90	49.54	67.67	97.51	97.51	14.51
Ht/Mat	33.42	19.18	13.24	93.72	95.07	1.35	21.58	22.37	23.37	99.64	24.65	34.13
Ht/MSW	35.90	59.01	4.17	53.14	93.33	48.64	92.93	6.19	0.00	90.32	16.64	19.25
Ht/GN	11.05	18.73	42.45	56.93	90.39	92.88	92.37	4.89	5.60	34.33	74.62	93.38
Ht/SPY	10.80	31.95	66.92	65.59	91.98	0.00	87.89	67.17	5.25	58.19	29.65	83.42

Table 7 Rank correlation coefficients between the predicted and observed cross rankings based on the phenotypic proportions in the F_3 samples

	$>P_1 > P_1$	$>P_1 < P_2$	$<P_2 > P_1$	$<P_2 < P_2$
Ht/TGW	0.150	0.550	0.750	0.900
Ht/MSW	0.750	0.850	0.750	0.900
Ht/MAT	0.800	0.950	0.800	0.900
Ht/AE	0.650	1.000	0.850	0.850
Ht/GN	0.900	0.700	0.800	0.925
Ht/SPY	0.225	0.900	0.625	0.900

based on the F_3 samples are given in table 7 where, with the possible exceptions of Ht/TGW and Ht/SPY in the $>P_1 > P_1$ category, the correspondence between observed and predicted ranking can be seen to be very good. The rank correlation coefficients based on the DH samples are given in table 8 and although they are better than the corresponding coefficients based on the information obtained from the genetical parameters they are inferior to the F_3 method.

Table 8 Rank correlation coefficients between the predicted and observed cross rankings based on the phenotypic proportions in the DH samples

	$>P_1 > P_1$	$>P_1 < P_2$	$<P_2 > P_1$	$<P_2 < P_2$
Ht/TGW	0.200	0.575	0.750	-0.050
Ht/MSW	0.425	0.775	0.750	0.225
Ht/Mat	0.375	0.125	0.800	0.225
Ht/AE	0.438	1.000	-0.200	0.900
Ht/GN	0.500	0.700	0.550	0.275
Ht/SPY	-0.375	-0.363	0.375	0.000

CONCLUSIONS

1. Multivariate predictions involving two, three and four variates have been made within crosses of spring barley. Although the sample size of 40 SSD lines was too small to allow statistical tests of the accuracy of the predictions they appear to provide reasonable agreement between observed and expected proportions.

2. For trivariate and higher order predictions the genetic correlations were replaced by phenotypic correlations and still resulted in accurate predictions.

3. The concept of cross prediction as a means of ranking crosses was further examined with reference to bivariate probabilities. The observed and predicted rankings from the F_3 and DH samples were compared by rank correlation procedures and in general the F_3 gave the closest agreement between observed and predicted ranking. Furthermore the predicted ranking of crosses was dependant on all three types of genetic components *i.e.*, means, variances and correlations.

4. An alternative method of ranking crosses was explored. Both F_3 and DH systems based on "progeny testing" gave better results than the corresponding methods based on genetical parameters. Nevertheless, the F_3 system was again superior to the DH method.

5. Although preliminary, the present data would suggest that "progeny testing" of crosses in barley would hold promise as a means of ranking crosses for combinations of characters of importance to barley breeders. An appropriate method could involve growing samples of F_3 families and then ranking the crosses on the basis of the observed proportion of lines exceeding given standards. Such a scheme would have the added attraction that first degree statistics, *i.e.*, means, would form the basis of the prediction technique.

Acknowledgements The advice and encouragement of Dr A. M. Hayter during the early stages of this work is gratefully acknowledged. We also acknowledge the technical assistance of Mr G. R. Drabble and Mr W. Wood.

REFERENCES

- CALIGARI, P. D. S., POWELL, W. AND JINKS, L. L. 1985. The use of doubled haploids in barley breeding. 2. An assessment of univariate cross prediction methods. *Heredity*, 353-358.

- OWEN, D. B. 1956. Tables for computing bivariate normal probabilities. *Annals of Mathematic Statistics*, 27, 1075-1090.
- POONI, H. S. AND JINKS, J. L. 1978. Predicting the properties of recombinant inbred lines derived by single seed descent for two or more characters simultaneously. *Heredity*, 40, 349-361.
- SCHERVISH, M. J. 1984. Multivariate normal probabilities with error bound. *Applied Statistics*, 33, 81-87.