Theor. Appl. Genet. 56, 225-232 (1980)

A Method of Parental Selection and Cross Prediction Using Incomplete Partial Diallels

Part 1: A Simulation Study

G.H. Gordon
Department of Agronomy, Waite Agricultural Research Institute, University of Adelaide, Adelaide (Australia)

Summary. When a character has a large additive genetic component in its variance, general combining ability estimates may be used for predicting cross performance. Further, if emphasis is placed on the ranks of the general combining abilities in a diallel rather than on their numerical values, the incomplete partial diallel is a powerful tool for parental selection. In a self pollinating species, if general combining ability effects are equal to or greater than specific combining ability effects, 20% of the partial diallel is found to give enough information for accurate ranking of the parental general combining abilities. A generalised method for calculating general combining abilities from incomplete partial diallels is presented and tested by computer simulation over a large range of genetic population parameters.

Key words: Diallel – Breeding – Selection – Prediction – Simulation – Progeny test

Introduction

A diallel is the set of all the possible crosses between a given number of parental lines. It can be represented by an N × N matrix where the ij^{th} element represents the hybrid between parents i and j and the leading diagonal (ii^{th}) elements represent the products of self-fertilization, that is, the parents.

The full diallel (F.D.) may conveniently be divided into three groupings: the parental lines; one set of crosses, where for each ij^{th} element $i < j$; and the reciprocal set of crosses where $i > j$.

A partial diallel is either the full diallel excluding the reciprocals or the full diallel excluding both the reciprocals and the parental lines.

In this paper, the term 'incomplete partial diallel' (I.P.D.) refers to any subset of the partial diallel.

Diallels have traditionally been used both to measure the mode of inheritance of quantitative genetic characters within a population and to isolate individual variations between parents in the diallel. A diallel experiment would normally have the observations of all the possible cross, and parental, genotypes carried out within a relatively uniform environment. This provides the mechanism for optimising the measurement of genetically controlled characters whilst helping to minimise environmental effects.

This genetic component of a quantitative character can be divided into three effects: the additive genetic effect of each of the parents involved in the cross [called general combining ability (gca)]; a specific effect caused by the interaction of the two parental genotypes [called specific combining ability (sca)]; and a reciprocal effect (r) due to the reversal of the male and female parents.

Thus, the model of a quantitative character (P) becomes:

$$P_{ij} = \text{mean effect} + \text{gca (parent i)} + \text{gca (parent j)} + \text{sca cress ij)} + r_{ij} \quad \text{(i)}$$

The model is completed by the addition of e_{ij} which represents the environmental effects as well as observational error.

Various methods have been proposed to estimate these effects from the diallel of observed values (Griffing 1956a, b; Dickinson and Jinks 1956; Hayman 1954a; Yates 1947). Such analyses may be of use to the geneticist and breeder for evaluating the mode of inheritance of a given character within a population.

With p parents used in a diallel, p(p-1) crosses have to be made and this rapidly becomes a very large number as p increases. For this reason, diallels with more than a few parents are rare and efforts are made to reduce their size. The simplest reduction is achieved by assuming a negligible reciprocal effect and thus eliminating reciprocal crosses. This reduces the number of crosses to p(p-1)/2

0040-5752/80/0056/0225/$ 1.60

and the model for the elements in this partial diallel becomes:

$$P_{ij} = \mu + gca_i + gca_j + sca_{ij} + e_{ij} \quad \text{(ii)}$$

One may predict the phenotype of a particular cross with the use of diallel-type analysis; once a full analysis of a diallel has been made, the right-hand side of Eq. (ii) may be used to estimate the left-hand side. While this may yield interesting information about the fit of the linear model, its practical applicability is questionable as the left-hand side values have already been observed. However, if all the crosses are not observed, the missing ones may be estimated as follows:

$$\hat{P}_{ij} = \bar{x} + \hat{gca}_i + \hat{gca}_j \quad \text{(iii)}$$

Here, a number of assumptions are made before such an estimate may fall within a reasonable range of accuracy. The first group of assumptions are the genetic assumptions of diallel analysis proposed by Hayman (1954b):

(a) diploid segregation,
(b) no difference between reciprocal crosses,
(c) independent action of non-allelic genes,
(d) no multiple alleles,
(e) homozygous parents,
(f) genes independently distributed between the parents.

As Hayman showed, assumption (d) is only important for generations after the F_1 generation. Gilbert (1958), Dickinson and Jinks (1956) and Kempthorne (1956) give a full discussion on the implications of these assumptions and methods of adjusting diallel analysis when some of them do not apply. Additional assumptions made here are:

(g) the environment effects within blocks are small;
(h) accurate estimates of the general combining ability can be made.

Use of a homozygous genotype as a control at predesigned placements in an experimental block allows for environmental effects to be evaluated, and correction for this variation allows acceptance of assumption (g).

Accepting that (g) is often true, there are two distinct options for estimating general combining ability [assumption (h)]:

1. the use of a partial diallel;
2. the use of an incomplete partial diallel.

Use of a partial diallel is, however, of little value, as the objective is to predict cross performance and the partial diallel involves all possible crosses. The partial diallel could, however, be used for estimating hybrids outside and/or between sets of diallels. Clearly it is not possible to estimate the result of a cross between a parent in a diallel and a parent which is not included in this diallel, as only the gca estimate of the parent in the diallel is available. If two diallels exist, and if one parent is in each diallel, then estimates of both gca's will be available and hence an estimate of the cross may be calculated.

However, diallel analysis with regard to gca estimating is an 'averaging' and 'fitting' process, involving all the hybrids produced from a parent and estimating their non-interactive genetic effect. This estimate is thus very sensitive to the set of genotypes used in the fitting process. Within a diallel, the gca estimates are fitted relative to other parents in the diallel. No information is necessarily obtained about parents outside the diallel nor about the effects of other parents on the diallel. For interdiallel crosses to be estimated with such gca's, it would be assumed that the addition and/or deletion of parents to either diallel would not affect the gca estimates of the remaining parental gca's. As Kempthorne (1956) showed such extrapolation from results within a diallel subset of a population to the population itself forms a weakness in the diallel method.

This leaves the incomplete partial diallel as a source of gca estimation. Here, all the gca's are estimated within the same set of genotypes. However, a similar question still remains, as to whether the addition (towards a full diallel) or deletion (towards a sparser incomplete partial diallel) of elements from the diallel affects the gca estimates of the parents. There will, however, always be a gca estimate for every parent that could be used in the predicted cross. Here the problem is more one of precision than of relativity.

Lupton (1965) discussed the value of estimates of gca from incomplete partial diallels for making predictions of hybrid performance. The incomplete partial diallel he used was the 'tester cross' method where each parent is crossed to the same set of tester varieties, with no crosses between testers. Although he concluded that this method has advantages over other methods he tested, he spent little time discussing the error inherent in the gca estimation of incomplete diallels compared to the partial diallel estimates. Fyfe and Gilbert (1963), Kempthorne and Curnow (1961), Curnow (1963) and Bray (1971) do, however, discuss the relative merits of various ways of sampling the diallel (incomplete partial diallel) as against using the partial diallel.

Ranking

These arguments regarding the accuracy of estimating combining ability effects are relevant when the objective in producing the diallel is to describe precisely the mode of inheritance – both for the population sampled and for individual members within that sample. As Gilbert (1958) points out, most breeders are only interested in the par-

ental lines and crosses themselves and not with their representativeness or otherwise of the population as a whole.

Consider the breeder's main problem: he/she initially has only two sets of information – one consists of the characteristics of existing genotypes, and the other consists of the desired characteristics of the genotypes he/she wishes to produce. Hence, the evaluation of existing and new parental material for hybridisation is an essential part of the breeder's task.

As shown earlier, before any prediction of unmade crosses can be used, the additive genetic effects (gca) must be estimated. These gca's have two estimable qualities: first, there is the gca value itself, usually expressed in units of the quantitative genetic character it represents; and secondly there is the relative positioning, or ranking, of parental gca's as compared to each other. Of the two, the latter is clearly more important to the breeder, who must work with the assumptions that two high-ranking gca individuals have greater breeding potential than two low-ranking gca individuals. Hence the standard error of the gca estimates will be largely ignored, for the remainder of this paper, in favour of the accuracy of gca ranking in the incomplete partial diallel, as compared to the partial diallel. If the ranking of gca values is seen as the critical measurement then methods of diallel analysis can be reassessed according to this different criterion.

Simulation Methods

A computer program was written, for the CDC Cyber 173, to simulate observations from a diallel of a self-pollinating species. The following specifications were used:

1. two variables GCAR and SCAR were chosen to represent the variances of the general and specific combining abilities respectively.

2. P values were chosen at random from a normal distribution with mean zero and variance GCAR, [N(0,GCAR)]. These values were then assigned as representing the gca values of the p parents in the diallel.

3. p(p+1)/2 values were then chosen at random from a normal distribution with mean zero and variance SCAR, [N(0,SCAR)]. These values were then assigned to represent the specific combining ability values plus the environmental effect of all the hybrids and selfs.

4. individual observations could then be generated using equation (iii), where the last two terms have been combined, i.e.

$$P_{ij} = \text{mean} + gca_i + gca_j + \underbrace{sca_{ij} + e_{ij}}_{sca_{ij}}$$

It should be noted that the values generated in steps 2 and 3 above do not conform exactly to the standard definition $\Sigma gca_i = \Sigma sca_{ij} = 0$. As they are taken at random from the normal distribution they will only approximate this restriction.

This simulated diallel can now be analysed as a complete or an incomplete partial diallel. It was systematically reduced to incomplete partial diallels representing 2, 4, 6, 8 and 10 crosses per parent. Each of these reduced diallels was then analysed for gca estimates which were ranked and compared to those of the partial diallel. The analysis took the form of a least square analysis to minimise (ΣE_{ij}^2) where

$$E_{ij} = P_{ij} - \hat{gca}_i - \hat{gca}_j - \bar{x}$$

The simulation was carried out for diallels of size P = 15, 20, 30 and 50. The number of diallels simulated for each of these parental numbers was 100, 50, 30 and 14 respectively. Each simulation replicate was independent of the others and represented a completely new set of parents. The number of replications was a limiting function of the computer time needed for the simulations. The random number generator used was the IMSL program library's GGNOR (IMSL 1977), using as its starting-point seed: SEED = (time of day) × (date), thus ensuring a unique starting-point for each simulation.

As the accuracy for the hybrid prediction using this method is clearly dependent on the proportional contribution of gca to sca, a number of GCAR and SCAR variables were used for the diallel generation. The values used were 1024, 256, 64, 16, 4 and 1. All permutations (36) of these values of the GCAR and SCAR values were used for each size of diallel. In all, 27,936 diallels were simulated. Figure 1 illustrates the method used for systematically sampling the partial diallel for the incomplete partial diallels.

The rank differences are evaluated by subtracting the rank obtained for a parent in the incomplete partial diallel from the rank of the same parent in the partial diallel. This is done for all parents and the absolute values of these differences are stored as ten values representing how many parents in the diallel varied in rank by zero, how many varied in rank by one and so forth, up to how many varied in rank by nine.

Results

a) Ranks

After each diallel was simulated and analysed, a tally was taken of the number of parents, in the incomplete partial diallels, which differed in rank from those in the partial diallel.

Of the 36 combinations of gca to sca variance ratios

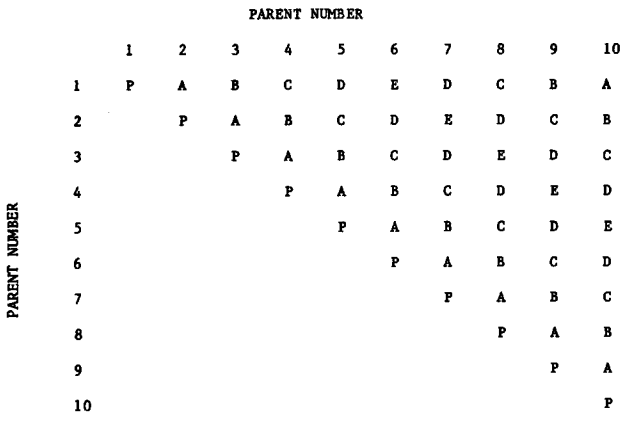

Fig. 1. This figure illustrates how the incomplete partial diallels were systematically sampled. The partial diallel consists of the parents (P) and all of the crosses (A, B, C, D and E). The two crosses per parent incomplete partial diallel consists of P + A. Similarly, the 4, 6 and 8 crosses per parent incomplete partial diallel consist of P + A + B, P + A + B + C, and P + A + B + C + D respectively

which were used, eleven were unique, e.g., the ratio of the variance of gca's from the distribution N(0,1024) to the variance of sca's from the N(0, 256) distribution is the same as for the two distributions N(0, 64) and N(0, 16) (1024/256 = 64/16 = 4.0). The eleven distinct ratios were 1024, 256, 64, 16, 4, 1, 0.25, 0.0625, 0.0156, 0.0039 and 0.00098 with replications 1, 2, 3, 4, 5, 6, 5, 4, 3, 2, and 1 respectively.

Duncan's Multiple Range Test (Duncan 1955) was performed on the components of the eleven groupings of variance ratios to see whether a scaling difference was present. No significant differences were found. Thus, the original thirty-six groupings of gca/sca ratios were reduced to eleven.

Of these remaining eleven groups, five had a larger gca component than sca component, one had equal components and five had smaller gca components than sca components. As would be expected, the latter five yielded many differences in the gca rank estimation. Table 1 shows Spearman's rank correlation coefficient for all the diallels simulated with a variance ratio of 1 or greater. The range and scatter of this coefficient did not differ appreciably for these ratios, so only the range (from the 256:1 ratio down to the 1:1 ratio) is shown. Although this table shows the stability of the ranks for the various incomplete partial diallels, it is difficult to visualise from this information alone the significance of the rank changes. For this reason, tables were also drawn up to show the number of gca estimates which did not change at all for the respective simulations. Similarly, tables showing rank changes of one or less, two or less and up to nine or less changes in rank were also compiled. As these tables are cumbersome,

Table 1. Range of Spearman's rank correlation coefficients for gca:sca variance ratios of 256:1 to 1:1. The ranks of the gca estimates from the partial diallel are compared to those from the incomplete partial diallel. The bracketed numbers represent the percentages of the crosses from the partial diallel used in the incomplete partial diallel

	15 parent diallel	20 parent diallel	30 parent diallel	50 parent diallel
10 crosses per parent I.P.D.	0.999-0.979 (71.4%)	0.999-0.968 (52.6%)	0.999-0.964 (34.5%)	0.999-0.966 (20.4%)
8 crosses per parent I.P.D.	0.999-0.969 (57.1%)	0.999-0.957 (41.2%)	0.999-0.955 (27.6%)	0.999-0.959 (16.2%)
6 crosses per parent I.P.D.	0.999-0.955 (42.9%)	0.999-0.950 (31.6%)	0.999-0.948 (20.7%)	0.999-0.952 (12.2%)
4 crosses per parent I.P.D.	0.999-0.940 (28.6%)	0.999-0.936 (21.1%)	0.999-0.935 (13.8%)	0.999-0.938 (8.2%)
2 crosses per parent I.P.D.	0.999-0.918 (14.3%)	0.999-0.896 (10.5%)	0.999-0.918 (6.9%)	0.999-0.912 (4.1%)

a subset of them appears here as Tables 2a, 2b and 2c showing, for a twenty parent diallel, the percentages of observed changes in the estimated gca ranks. The complete set of such tables is available from the author. Here the full range of variance ratios is shown to indicate the rapid loss of information which occurs when the gca/sca ratio drops below one.

From these tables it is possible to estimate the probability of the breeder making an incorrect decision based on an incomplete partial diallel. He may, for example, be looking at 20 possible parental lines and wish to choose those with the top five gca values. If a parent is ranked fifth then any rank change towards an increased rank will cause it to be incorrectly ranked outside the top five. From Table 2a (in a twenty-parent diallel with two crosses per parent and a variance ratio of 64:1 (100-77.87)% of the diallel would be expected to change rank. As the rank change is only important in one direction, this statistic can be halved to encompass only changes towards the lower ranks. The approximate chance of the fifth ranked parent not being ranked in the top five then becomes:

$$\tfrac{1}{2}(100-77.87)\% = \tfrac{1}{2}(22.13)\%$$
$$= 11.07\%$$

Similarly, from Tables 2b and 2c the estimates for the fourth and third ranked parents become:

½(100–97.4)% = 1.35% (Table 2b)

½(100–99.67)% = 0.165% (Table 2c)

The probability for the first and second ranked parent are obtained from the rank difference tables of four or less and three or less respectively (not presented here).

Table 3 summarises these results for the 15, 20, 30 and 50 parental diallels for gca/sca ratios $\geqslant 1$, where the incomplete partial diallel contains approximately 20% of all possible crosses. The 20% (or its closest approximate) incomplete partial diallel was chosen, as this was observed to be a reasonable cut-off point between having an excessively large number of crosses and having too little information. The selection of five parents only is used, as this would be a practical number for a plant breeder to use in more extensive experimentation. It is however, only an example and plant breeders may in fact wish to choose more (or fewer) parents from their original diallel.

As the probabilities in these tables are associated with the probabilities of all of the parents and not just the first five ranked parents, these estimates can be considered upper limits and accordingly it would be expected that the exact probabilities would be lower. For this reason and for the sake of brevity, standard errors associated with these estimates were not included in Table 3.

Table 2a. Means and standard errors of the percentage of parents in 20 x 20 diallel whose GCA estimates do not change in rank when the incomplete partial diallel analysis is compared to the partial diallel analysis

GCA/SCR variance ratio	2 crosses per parent (20/190=10.5% of possible crosses)	4 crosses per parent (40/190=21.1% of possible crosses)	6 crosses per parent (60/190=31.6% of possible crosses)
1024.0	91.00 ± 1.12	93.40 ± 1.05	94.00 ± 1.03
256.0	87.30 ± 0.99	88.60 ± 1.03	90.90 ± 0.97
64.0	77.87 ± 1.03	81.27 ± 0.99	85.23 ± 0.90
16.0	62.82 ± 0.91	68.45 ± 0.99	72.30 ± 0.87
4.0	43.54 ± 0.81	49.48 ± 0.87	55.56 ± 0.93
1.0	25.67 ± 0.64	29.95 ± 0.69	36.22 ± 0.70
0.2500	15.78 ± 0.58	19.62 ± 0.57	22.24 ± 0.66
0.0625	11.95 ± 0.52	13.87 ± 0.55	15.20 ± 0.63
0.0156	9.00 ± 0.53	10.33 ± 0.55	11.80 ± 0.68
0.0039	8.95 ± 0.64	9.90 ± 0.72	12.10 ± 0.80
0.0010	9.60 ± 0.99	10.90 ± 1.00	10.30 ± 0.84

Table 2b. Means and standard errors of the percentage of parents in a 20 x 20 diallel whose GCA estimates do not change in rank more than one when the incomplete partial diallel analysis is compared to the partial diallel analysis

GCA/SCR variance ratio	2 crosses per parent (20/190=10.5% of possible crosses)	4 crosses per parent (40/190=21.1% of possible crosses)	6 crosses per parent (60/190=31.6% of possible crosses)
1024.0	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
256.0	99.10 ± 0.26	99.20 ± 0.24	99.50 ± 0.19
64.0	97.30 ± 0.37	97.70 ± 0.35	98.23 ± 0.33
16.0	90.78 ± 0.59	93.15 ± 0.51	94.97 ± 0.43
4.0	77.84 ± 0.71	83.00 ± 0.69	86.94 ± 0.59
1.0	56.67 ± 0.75	63.02 ± 0.75	69.00 ± 0.66
0.2500	38.82 ± 0.81	45.16 ± 0.75	49.04 ± 0.77
0.0625	29.00 ± 0.73	34.55 ± 0.78	36.82 ± 0.82
0.0156	25.53 ± 0.76	26.37 ± 0.82	29.40 ± 0.88
0.0039	24.40 ± 1.13	25.30 ± 0.97	31.80 ± 1.13
0.0010	25.10 ± 1.52	24.60 ± 1.55	28.30 ± 1.49

Table 2c. Means and standard errors of the percentage of parents in a 20 x 20 diallel whose GCA estimates do not change in rank more than two when the incomplete pzrtial diallel analysis is compared to the partial diallel analysis

GCA/SCR variance ratio	2 crosses per parent (20/190=10.5% of possible crosses)	4 crosses per parent (40/190=21.2% of possible crosses)	6 crosses per parent (60/190=31.6% of possible crosses)
1024.0	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
256.0	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
64.0	99.67 ± 0.15	99.80 ± 0.09	99.83 ± 0.09
16.0	97.82 ± 0.29	98.73 ± 0.20	99.25 ± 0.17
4.0	92.58 ± 0.41	94.92 ± 0.39	96.38 ± 0.34
1.0	74.38 ± 0.70	80.52 ± 0.60	85.18 ± 0.55
0.2500	55.24 ± 0.80	62.42 ± 0.74	68.10 ± 0.79
0.0625	43.97 ± 0.85	49.40 ± 0.81	53.40 ± 0.85
0.0156	38.70 ± 0.91	40.83 ± 0.95	44.77 ± 1.00
0.0039	35.50 ± 1.22	38.60 ± 0.99	45.05 ± 1.13
0.0010	36.90 ± 1.50	38.40 ± 1.62	42.50 ± 1.38

b) Prediction

The values of the gca estimates were used to 'predict' the partial diallel components using Eq. (iii). Each of these estimated crosses can then be compared to the observed value (in this case the original simulated value). The absolute value of the expected value less the observed value was calculated for each element of the diallel, and then a mean $\left(\frac{\Sigma|\text{Exp} - \text{Obs}|}{N}\right)$ determined for the entire partial diallel. The calculations were done with all the incomplete partial diallels that were tested, and these results are summarised in Table 4. The absolute value was used as it was felt that the magnitude rather than the direction of the deviation from the observed was the important statistic.

Table 3. Probability of the top five ranked parents being incorrectly ranked outside of the top five when the incomplete partial diallel is compared to the partial diallel

	Ranking	Variance ratio					
		1024	256	64	16	4	1
15 parent diallel	1	0.0000	0.0000	0.0000	0.0001	0.0005	0.0062
4 crosses/parent	2	0.0000	0.0000	0.0000	0.0005	0.0023	0.0191
(30/105=28% of	3	0.0000	0.0000	0.0003	0.0027	0.0127	0.0501
possible crosses)	4	0.0000	0.0017	0.0060	0.0224	0.0495	0.1274
	5	0.0207	0.0330	0.0721	0.1248	0.2013	0.3063
20 parent diallel	1	0.0000	0.0000	0.0000	0.0004	0.0017	0.0217
4 crosses/parent	2	0.0000	0.0000	0.0002	0.0021	0.0069	0.0488
(40/190=21.1% of	3	0.0000	0.0000	0.0010	0.0064	0.0254	0.0974
possible crosses)	4	0.0000	0.0040	0.0115	0.0342	0.0850	0.1849
	5	0.0030	0.0570	0.0937	0.1577	0.2526	0.3502
30 parent diallel	1	0.0000	0.0000	0.0002	0.0004	0.0064	0.0419
6 crosses/parent	2	0.0000	0.0000	0.0002	0.0026	0.0207	0.0790
(90/435=20.7% of	3	0.0000	0.0006	0.0020	0.0138	0.0530	0.1424
possible crosses)	4	0.0000	0.0061	0.0135	0.0589	0.1329	0.2414
	5	0.0444	0.0678	0.1259	0.2244	0.3167	0.3906
50 parent diallel	1	0.0000	0.0000	0.0000	0.0021	0.0214	0.0973
10 crosses/parent	2	0.0000	0.0000	0.0010	0.0082	0.0520	0.1487
(250/1225=20.4% of	3	0.0000	0.0011	0.0031	0.0325	0.1026	0.2196
possible crosses)	4	0.0029	0.0064	0.0329	0.0991	0.2011	0.3096
	5	0.0414	0.0996	0.1707	0.2771	0.3693	0.4230

Table 4. Means and standard errors of the mean deviation of the predicted partial diallel and the incomplete partial diallel. All values are based on a gca:sca variance ratio of 1024:64. The bracketed numbers represent the percentage of the crosses from the partial diallel used in the incomplete partial diallel

	15 parent diallel	20 parent diallel	30 parent diallel	50 parent diallel
Partial Diallel	6.04 ± 0.45 (100.0%)	6.03 ± 0.65 (100.0%)	1.91 ± 0.21 (100.0%)	1.94 ± 0.31 (100.0%)
10 crosses per parent I.P.D.	6.14 ± 0.46 (71.4%)	6.19 ± 0.67 (52.6%)	1.98 ± 0.23 (34.5%)	2.01 ± 0.34 (20.4%)
8 crosses per parent I.P.D.	6.22 ± 0.47 (57.1%)	6.28 ± 0.68 (42.1%)	2.00 ± 0.23 (27.6%)	2.03 ± 0.35 (16.2%)
6 crosses per parent I.P.D.	6.33 ± 0.48 (42.9%)	6.37 ± 0.70 (31.6%)	2.02 ± 0.24 (20.7%)	2.06 ± 0.36 (12.2%)
4 crosses per parent I.P.D.	6.50 ± 0.50 (28.6%)	6.51 ± 0.72 (21.2%)	2.05 ± 0.25 (13.8%)	2.19 ± 0.37 (8.2%)
2 crosses per parent I.P.D.	6.71 ± 0.52 (14.3%)	6.77 ± 0.76 (10.5%)	2.11 ± 0.27 (6.9%)	2.16 ± 0.40 (4.1%)

For illustrative purposes only the gca/sca variance ratios of 1024:64 are shown in Table 4. As the mean deviation is related to the sca variance, the complete range of variance ratios may be calculated, from this table, by use of the following formula:

$$\text{new mean deviation} = \text{old mean deviation} \times \sqrt{\text{new sca variance}} \div \sqrt{64}$$

[For a 15 parent diallel with 4 crosses for parent, the mean deviation expected with a gca:sca variance ratio of 256:16 would be:

$$6.04 \times \sqrt{16} \div 8 = 3.02]$$

Discussion

As can be seen from Tables 1, 2a, 2b and 2c very little change occurs in the ranking of the gca estimates when the partial diallel is reduced.

As would be expected, the gca/sca variance ratios which have a higher sca component than gca component lead to inaccuracies in the incomplete partial diallel rank-

ing. This indicates that only characters with a gca/sca ratio of one or greater should be analysed in an incomplete partial diallel. Whilst this restriction excludes some characters, many agronomic characters conform to this restriction (cf. Griffing 1956b; Reddy 1976; Chaudhary et al. 1977; Dhillon and Singh 1977; and Lupton 1965).

The added ranking information gained by using the partial diallel would not outweigh the extra work involved, instead of making a partial diallel with twenty parents (190 crosses) twenty percent of a forty parent partial diallel may be made (160 crosses). Here, fewer crosses have been made but twice the number of parents have been included. In terms of finding the best parents to use in a breeding program, little information is lost by using only a fraction of the diallel. Table 3 shows the estimated upper probabilities for parentals being ranked outside the top five; for most gca/sca variances the probability of not finding the top three parentals is small. If more than five parents are to be chosen, these probabilities become even smaller.

A critical part of this analysis is the inclusion of the parental lines themselves. Simulations were run excluding the parents and it was found that the number of observed values had to remain the same for similar results to be obtained. That is, if the twenty parents were not used in the diallel then twenty additional crosses would have to replace them in the analysis to give results with approximately the same number of ranking changes. As increasing the number of crosses necessary is disadvantageous, the parents were used in the diallel instead.

On a theoretical basis, this can be justified on the grounds that the plant breeder is fundamentally only interested in the breeding material itself and thus any bias introduced by the inclusion of the parental lines is a bias within the population genetic parameters of the species, rather than a bias within the genotypic subset used. As the breeder is mainly concerned with the bias affecting the parental lines he is selecting from, bias that affects information on parents outside the diallel is tolerable.

The simulated diallels were in the form of mean values for each genotype; in the field, the observations would of course be in the form of replicates of some convenient number. As replicated trials are the more usual use of least squares analysis, the method is easily adapted to multiple replicate input rather than input in the form of means. Because of the way in which the mean values were simulated, input in the form of replicated trials was not necessary for the simulation testing of the method. Also, as the diallel consists only of parental lines and F_1's, replication is only needed to minimise the environment differences, which are not relevant to this type of computer simulation.

Rather than looking at the best parental gca, all possible crosses between parents can be predicted [Eq. (iii)] and then the best of the predicted crosses chosen. With lines appraised for a number of characters, the use of cross prediction would give a simpler selection procedure to the breeder than the individual use of the gca estimates for each character.

In terms of the accuracy of these predicted crosses as compared to the observed values, Table 4 gives an indication of the critical features of this simulation. The most important point to note from the table is the closeness of fit between the predictions based on the partial diallel and those based on the incomplete partial diallel. It is also observed that as the diallel increases in size the mean deviation decreases, which seems to be due to the larger sample size (same number of crosses per parent – but more parents) having a 'buffering' effect on the |Obs-Exp| values. Using Tables 3 and 4, the following breeding plan is suggested:

Generation 1

(1) A decision is made by the breeder on the parents He/she wishes to assess.

(2) The parental lines are grown and crosses made to give approximately twenty percent of the partial diallel crosses, as suggested in this paper.

Generation 2

(3) The incomplete partial diallel is observed for the quantitative genetic character considered important.

(4) The least squares diallel analysis, as suggested in this paper, is undertaken on the observed values obtained in 3 above.

(5) Using the average error of the observed less the predicted values of the diallel observations and the variance of the estimated gca values, the gca/sca ratio may be approximated from Table 4. Table 3 can be used to approximate the probability of choosing the top five parents. If this probability is not favourable then more than five parental lines may have to be chosen (similarly fewer may be chosen).

Generation

(6) Embark on normal breeding program with parental lines chosen. For other than half seasonal crops, this plan would have to be extended a year as the diallel would now have to be grown in a different year to part 1.

As the accuracy of the gca parental ranking procedure depends critically on the gca/sca variance ratio, step five is vitally important and yet in a sense irrelevant. If the gca/sca variance is unknown, so is the precision of the parental selection, however, short of increasing the size of the incomplete partial diallel in a breeding program, this is the most accurate information attainable. If the accuracy is suspect, then the number of parents chosen may have to be increased, but provided it is not increased to include all the parental material, some enhancement of efficiency has been obtained.

Step five may not be necessary if the breeder has an

approximation of the gca/sca ratio for that character in the population, either from the literature or from his/her own diallel analysis of a subset of the parental lines. As stated earlier, this extrapolation from a diallel subset to the population may not be precise but it should give a reasonable approximation to the population gca/sca variance ratio.

The most important aspect of this method is not any ability to select the parental lines precisely, but rather to help the breeder increase the chances of finding these lines. Like all breeding methods this one yields no certain way of finding the one individual in the many thousands that the breeder is seeking, but any methods for a given input of resources, which improve the chances of isolating a genotype that approaches the optimum, are worth implementing.

The breeder is limited in the number of plants that can be managed; with many crosses early generation selection must be much more intense than if only a few crosses were used. Thus, by minimising the number of crosses, a greater number of the more relevant later generation selections may be grown, giving more chance of isolating the genetic extremes sought. The postulate that F_1 prediction is correlated with later generation homozygotes is beyond the scope of this paper, however, F_1 prediction has direct use for hybrid crops.

It should be noted that although variances and standard errors associated with the gca estimations have been largely ignored throughout this paper, the least squares analysis method used does provide this type of information. However, as shown earlier, this is really only of importance in the partial diallel case and thus the particular form of analysis used is also suitable for conventional diallel analysis. For purposes of testing, all the least square diallel analyses were done in parallel with the standard method of Griffing (1956b). For the partial diallel case both methods gave identical results, but for the incomplete partial designs, the two methods gave markedly different results, the overall ranking remaining far more consistent with the generalised least square method. The least squares computer program used is available from the author both in the form of a PASCAL language program as well as in flowchart form, suitable for translation into FORTRAN.

Acknowledgement

The author wishes to thank Professor C.J. Driscoll and Dr O. Mayo for their valuable advice and encouragement.

Literature

Bray, R.A.: Quantitative evaluation of the circulant partial diallel cross. Heredity 27, 189-202 (1971)

Chaudhary, B.D.; Kakar, S.N.; Singh, R.K.: Estimation of genetic parameters in barley. II. Partial diallel analysis. Theor. Appl. Genet. **49**, 153-156 (1977)

Curnow, R.N.: Sampling the diallel cross. Biometrics **19**, 287-306 (1963)

Dhillon, B.S.; Singh, J.: Combining ability and heterosis in diallel crosses of maize. Theor. Appl. Genet. **49**, 117-122 (1977)

Dickinson, A.G.; Jinks, J.L.: A generalised analysis of diallel crosses. Genetics **41**, 65-78 (1956)

Duncan, D.B.: Multiple range and Multiple F tests. Biometrics **11**, 1-42 (1955)

Fyfe, J.L.; Gilbert, N.: Partial diallel crosses. Biometrics **19**, 278-286 (1963)

Gilbert, N.E.G.: Diallel crosses in plant breeding. Heredity **12**, 477-492 (1958)

Griffing, B.: A generalised treatment of the use of diallel crosses in quantitative inheritance. Heredity **10**, 31-50 (1956a)

Griffing, B.: Concept of general and specific combining ability in relation to diallel crossing systems. Austr. J. Biol. Sci. **9**, 463-493 (1956b)

Hayman, B.I.: The analysis of variance of diallel tables. Biometrics **10**, 235-244 (1954a)

Hayman, B.I.: The theory and analysis of diallel crosses. Genetics **39**, 789-809 (1954b)

I.M.S.L.: Houston, Texas, U.S.A.: International Mathematical and Statistical Libraries Inc. 1977

Kempthorne, O.: The theory of the diallel cross. Genetics **41**, 451-459 (1956)

Kempthorne, O.; Curnow, R.N.: The partial diallel cross. Biometrics **17**, 229-250 (1961)

Lupton, F.G.H.: Studies in the breeding of self pollinating cereals. 5. Use of the incomplete diallel in wheat breeding. Euphytica **14**, 331-352 (1965)

Reddy, L.V.: Combining ability analysis of some quantitative characters in hexaploid triticale. Theor. Appl. Genet. **47**, 227-230 (1976)

Yates, F.: Analysis of data from all possible reciprocal crosses between a set of parental lines. Heredity **1**, 287-301 (1947)

Received June 8, 1979
Communicated by J.S.F. Barker and W.J. Libby

Dr. G.H. Gordon
Department of Genetics
University of Birmingham
P.O. Box 363
Birmingham B15 2TT (England)