Theor. Appl. Genet. 55, 129-132 (1979)

THEORETICAL AND APPLIED GENETICS

Doubled Haploids for Estimating Genetic Variances in Presence of Linkage and Gene Association

T.M. Choo and E. Reinbergs
Department of Crop Science University of Guelph, Guelph, Ontario (Canada)

Summary. The effect of gene association (or dispersion) and linkage on the estimation of genetic variances in a diallel experiment involving doubled haploid lines is evaluated. It is shown that the estimates of the additive and the additive × additive genetic variances, as obtained by Choo et al. (1979), are biased if genes are linked or are not independently distributed in the parents. However, this bias only occurs in the presence of interaction between homozygous loci. Gene association (or dispersion) and linkage, if present, can be detected by comparing the parental vs the crosses mean, the parental vs the doubled haploid lines variance, and the among vs the within crosses variance.

Key words: Diallel analysis – Gene dispersion – Self pollinating crops – Linkage disequilibrium

---

Introduction

The procedure for obtaining doubled haploids in barley have been described by Kasha and Kao (1970), Subrahmanyam and Kasha (1975) and Jensen (1975). Using doubled haploid techniques in barley breeding, truly homozygous lines can be obtained from the gametes of F_1 hybrids and thereby, the time required to develop new cultivars may be lessened by at least three generations (Kasha and Reinbergs 1975). In terms of grain yield, the results from the doubled haploid method have been similar to those obtained by the pedigree and single seed descent methods (Park et al. 1976) and the bulk method (Song et al. 1978).

With the recent development of doubled haploid and the single seed descent (Brim 1966) methods, population improvement programs, which have been shown to be an effective approach for cross-pollinating crops, are now feasible in self-pollinating crops. Such population improvement programs are being conducted in oats using the single seed descent method (D.D. Stuthman, personal communication) and in barley using the doubled haploid procedures at this institution. Choo et al. (1979) proposed a scheme for population improvement of self-pollinating crops. In this scheme, inbred parents are crossed in a diallel series and then doubled haploid (DH) homozygous lines (or near-homozygous lines if single seed descent method is used) are derived from each F_1 family. (F_1 families hereafter will be referred to as 'crosses'.) The system relative to gene differences and recombination (r) may be illustrated as follows:

$$\begin{array}{ll}
\text{Parents} & \mathrm{AABB} \times \mathrm{aabb} \\
\downarrow & \\
F_1 & \mathrm{AB/ab} \\
\downarrow & \\
\text{haplotypes} & \frac{1-r}{2}\mathrm{AB} + \frac{1-r}{2}\mathrm{ab} + \frac{r}{2}\mathrm{Ab} + \frac{r}{2}\mathrm{aB} \\
\downarrow & \\
\text{DH progeny} & \frac{1-r}{2}\mathrm{AABB} + \frac{1-r}{2}\mathrm{aabb} + \frac{r}{2}\mathrm{AAbb} + \\
 & + \frac{r}{2}\mathrm{aaBB}
\end{array}$$

The homozygous lines are evaluated in hill plots and the superior lines are selected as parents for the next cycle.

To estimate genetic variability in a doubled haploid population Choo et al. (1979) developed a statistical analysis for a diallel experiment in which each cross is represented by a number of doubled haploid lines. They showed that the variance of parents ($\sigma^2{}_p$) is the sum of the additive ($\sigma^2{}_A$) and the additive × additive ($\sigma^2{}_{AA}$) genetic variances whereas the variance among crosses ($\sigma^2{}_c$) is one-half of the additive variance plus one quarter of the additive × additive variance. Thus both additive and additive × additive genetic variances can be obtained by solving the above two equations; for example, $\sigma^2{}_{AA} =$

040-5752/79/0055/0129/$ 1.00

$2(\sigma_p^2-\sigma_c^2)$. However, this formulation is based upon two basic assumptions, namely, independent distribution of genes in the parents and no linkage. It is conceivable that there are deviations from these two assumptions in many diallel experiments. The objective of this study was to investigate the effect of gene association (or dispersion) and linkage on the estimation of genetic variances in a diallel experiment involving doubled haploid lines.

Non-Independent Gene Distribution and Linkage

If genes are independently distributed in the parents, the frequencies of inbred parents, AABB, AAbb, aaBB, aabb, can be assumed as U_aU_b, U_aV_b, V_aU_b, V_aV_b, respectively. The genotypic values, measured as deviations from the mid-parent value are $+d_a$, $-d_a$ for AA, aa, and $+d_b$, $-d_b$ for BB, bb respectively. However, in many instances, genes in the parents are not independently distributed. Gene association and dispersion may result from linkage of genes in the population from which the parents were chosen or from the effects of sample size when sampling the parents. Therefore, in order to describe gene association or dispersion, the parameter, 'c' is used in this study (See Coughtrey and Mather 1970; Mather and Jinks 1971). When c is positive the parents with reinforcing combinations of genes are in excess and the genes may be said to be predominantly associated. On the other hand, a negative c indicates that the parents with balancing combinations of genes are in excess, in this case, the genes may be said to be predominantly dispersed. If the distributions of the genes in the parents are independent of one another, c is equal to zero.

Let us consider that the genes are linked and are not independently distributed in the parents. We use c to describe gene association and dispersion, and also use r and i to denote the probability of recombination and the interaction between the two loci when both are homozygous, respectively. We assume that inbred parents are crossed in a diallel series and a number of doubled haploid lines are derived from each different F_1 family. Accordingly, the frequencies and means of the doubled haploid progenies from different F_1 families can be calculated (Table 1). It can be shown that the mean of parents is:

$$\bar{X}_p = (U_a - V_a)\,d_a + (U_b - V_b)\,d_b + [(U_a - V_a)(U_b - V_b) + 4c]\,i$$

Table 1. Phenotypes and means of parents and doubled haploid progenies of a diallel cross involving two gene pairs with interaction, association (or dispersion), and linkage

	Parents (♂)	AABB	AAbb	aaBB	aabb
	Phenotype	d_a+d_b+i	d_a-d_b-i	$-d_a+d_b-i$	$-d_a-d_b+i$
	Frequency	U_aU_b+c	U_aV_b-c	V_aU_b-c	V_aV_b+c
Parent (♀)	AABB	(AABB)*	(½AABB,½AAbb)	(½AABB,½aaBB)	+
Phenotype	d_a+d_b+i	d_a+d_b+i**	d_a	d_b	$(1-2r)i$
Frequency	(U_aU_b+c)	$(U_aU_b+c)^2$ ***	$(U_aU_b+c)\,(U_aV_b-c)$	$(U_aU_b+c)\,(V_aU_b-c)$	$(U_aU_b+c)\,(V_aV_b+c)$
Parent (♀)	AAbb	(½AABB,½AAbb)	(AAbb)	++	(½AAbb,½aabb)
Phenotype	d_a-d_b-i	d_a	d_a-d_b-i	$(2r-1)i$	$-d_b$
Frequency	(U_aV_b-c)	$(U_aV_b-c)\,(U_aU_b+c)$	$(U_aV_b-c)^2$	$(U_aV_b-c)\,(V_aU_b-c)$	$(U_aV_b-c)\,(V_aV_b+c)$
Parent (♀)	aaBB	(½AABB,½aaBB)	++	(aaBB)	(½aaBB,½aabb)
Phenotype	$-d_a+d_b-i$	d_b	$(2r-1)i$	$-d_a+d_b-i$	$-d_a$
Frequency	(V_aU_b-c)	$(V_aU_b-c)\,(U_aU_b+c)$	$(V_aU_b-c)\,(U_aV_b-c)$	$(V_aU_b-c)^2$	$(V_aU_b-c)\,(V_aV_b+c)$
Parent (♀)	aabb	+	(½AAbb,½aabb)	(½aaBB,½aabb)	(aabb)
Phenotype	$-d_a-d_b+i$	$(1-2r)i$	$-d_b$	$-d_a$	$-d_a-d_b+i$
Frequency	(V_aV_b+c)	$(V_aV_b+c)\,(U_aU_b+c)$	$(U_aV_b+c)\,(U_aV_b-c)$	$(V_aV_b+c)\,(V_aU_b-c)$	$(V_aV_b+c)^2$

* Phenotype(s) of the doubled haploid progenies derived from the F_1 family of AABB×AABB

** Mean phenotypic value of all doubled haploid progenies derived from the F_1 family of AABB×AABB

*** Frequency of all doubled haploid progenies derived from the F_1 family of AABB×AABB

+ $(\frac{1-r}{2}$ AABB, $\frac{1-r}{2}$ aabb, $\frac{r}{2}$ AAbb, $\frac{r}{2}$aaBB)

++ $(\frac{r}{2}$AABB, $\frac{r}{2}$ aabb, $\frac{1-r}{2}$ AAbb, $\frac{1-r}{2}$ aaBB) whereas the mean of crosses (doubled haploids from different F_1s) is:

$$\begin{aligned}\bar{X}_c &= (U_a - V_a)\,d_a + (U_b - V_b)\,d_b \\ &\quad + [(U_a - V_a)(U_b - V_b) + 2c + 2c\,(1 - 2r)]\,i\end{aligned}$$

Thus, for genes that are linked and are not independently distributed in the parents, the mean of crosses would not be the same as the parental mean except when genes are tightly linked ($r = 0$). The variance of parents is:

$$\begin{aligned}\sigma_p^2 &= 4\,U_aV_a\,[d_a + (U_b - V_b)i]^2 + 4\,U_bV_b\,[d_b \\ &\quad + (U_b - V_b)i]^2 + 16\,U_aV_aU_bV_b\,i^2 \\ &\quad + 8c\,d_ad_b - 8\,c\,[(U_a - V_a)i\,d_a + (U_b - V_b)i\,d_b] \\ &\quad - 8c\,(U_a - V_a)(U_b - V_b)i^2 - 16\,c^2i^2,\end{aligned}$$

which differs from the variance among doubled haploid lines, namely:

$$\begin{aligned}\sigma_{DH}^2 &= 4\,U_aV_a\,[d_a + (U_b - V_b)i]^2 + 4\,U_bV_b\,[d_b \\ &\quad + (U_b - V_b)i]^2 + 16\,U_aV_aU_bV_b\,i^2 \\ &\quad + 8c\,(1 - r)\,d_ad_b - 8c\,(1 - r)\,[(U_a - V_a)i\,d_a \\ &\quad + (U_b - V_b)i\,d_b] - 8c\,(1 - r)\,(U_a - V_a)(U_b - V_b)i^2 \\ &\quad - 16\,(1 - r)^2\,c^2i^2\end{aligned}$$

except when genes are tightly linked ($r = o$).
The variance among crosses,

$$\begin{aligned}\sigma_c^2 &= 2\,U_aV_a\,[d_a + (U_b - V_b)i]^2 + 2\,U_bV_b\,[d_b \\ &\quad + (U_a - V_a)i]^2 + 8\,(1 - 2r + 2r^2)\,U_aV_aU_bV_bi^2 \\ &\quad + 4c\,d_ad_b \\ &\quad - 4\,c\,(1 - 2r)\,[(U_a - V_a)id_a \\ &\quad + (U_b - V_b)i\,d_b] - 4c\,(1 - 2r^2)(U_a - V_a)(U_b - V_b)i^2 \\ &\quad - 8c^2\,(1 - 2r)i^2,\end{aligned}$$

is very unlikely to be similar to the variance within crosses,

$$\begin{aligned}\sigma_w^2 &= 2\,U_aV_a\,[d_a + (U_b - V_b)i]^2 + 2\,U_bV_b\,[d_b \\ &\quad + (U_a - V_a)i]^2 + 8\,(1 + 2r - 2r^2)\,U_aV_aU_bV_b\,i^2 \\ &\quad + 4c\,(1 - 2r)\,d_ad_b \\ &\quad - 4\,c\,[(U_a - V_a)\,i\,d_a + (U_b - V_b)\,i\,d_b] \\ &\quad - 4\,c\,(1 - 2r + 2r^2)(U_a - V_a)(U_b - V_b)i^2 \\ &\quad - 8c^2\,(1 - 2r + 2r^2)i^2\end{aligned}$$

except at $r = o$.

Gene association and dispersion may arise from the sampling error even though the loci are not linked, i.e. $r = 0.5$. In this case, r is substituted by 0.5 in the above equations. Obviously, the parental mean is $2ci$ greater and smaller than the crosses mean under gene association and gene dispersion conditions, respectively. The variance of the parents again is not equal to the variance among doubled haploid lines. Also, the variances among crosses and within crosses are very unlikely to be homogenous.

For genes that are linked but are independently distributed in the parents, i.e. $c = 0$, the parents and the crosses have equal means, and the variance of parents is the same as the variance among doubled haploid lines. These relationships are unlike to non-zero c conditions. When interaction between homozygous loci is present, i.e. $i \neq 0$, the variance among crosses in the case of linkage is $4\,U_aV_aU_bV_b\,(1-2r)^2\,i^2$ greater than that under no linkage conditions. On the other hand, the variance within crosses is $4\,U_aV_aU_bV_b\,(1-2r)^2\,i^2$ smaller. The amount of increase and decrease of the variance depends upon the probability of recombination. If the two loci are closely linked (i.e., r near zero), the variances among and within crosses are identical or similar. Thus, by comparing the variances among and within crosses, it can be determined if the loci are tightly linked.

Discussion

The formulae for the estimation of additive × additive genetic variance, suggested previously by Choo et al. (1979), are based upon the assumptions that the genes are independently distributed and that linkage is absent. If interaction between homozygous loci is not present in the research material, then estimates of additive gene action will be unbiased, regardless of the presence or absence of the above two conditions.

When there is interaction between homozygous loci ($i \neq 0$), the estimates of genetic variance can be biased unless the genes are independently distributed and are not linked. The biases may be upward or downward depending upon whether genes are associated or dispersed. The extent of the bias is determined by both the degree of gene association (or dispersion) and the probability of recombination. This should be noted when interpreting the results of a diallel experiment. However, the failure of one or both of the above assumptions can be detected by comparing the parental mean vs the crosses mean, the parental variance vs the doubled haploid variance, and the among crosses vs the within crosses variance. Genes are believed to be independently distributed if the parents and the crosses have equal means and also the parents and the doubled haploid lines have homogenous variances. Otherwise, they may not be independently distributed.

Whether genes are associated or dispersed can be determined by examining the difference between the parental and crosses means. The loci could be tightly linked together when the variance among crosses is found to be similar to the variance within crosses. If the following relationships are all true: (1) the parents and the crosses have equal means; (2) the parents and the doubled haploid lines have homogenous variances; and (3) both variances among and within crosses are one-half of the variance of parents, then, the additive × additive genetic variance is not significant.

We have used a two-locus system to investigate the effect of linkage and gene association (or dispersion) on the estimation of genetic variances, in a diallel experiment involving doubled haploid lines. In the case of a multiple-locus system, the effect of linkage and gene association between one pair of genes could be counteracted by that of other pairs and thus, only the net effect could be detected.

Based on the results of Choo et al. (1979) and this report, it appears that a diallel experiment involving doubled haploid lines can be used not only to estimate additive and additive × additive genetic variances but also to test for additive epistasis, linkage and gene association. It should be noted that if testing for linkage and gene association are the main objectives of a diallel experiment, then the experiment should have sufficient accuracy to detect the effect of linkage and gene association. On the other hand, if the main interest is to estimate the genetic variances in the experimental material, the same level of accuracy may not be required because, if the effect of linkage and gene association is too small to detect, it can be assumed that the biases of the estimation of genetic variances are insignificant.

In a recent review article on analytical methods for diallel experiments in which F_1 or F_2 bulks are used for evaluation, Baker (1978) pointed out that (1) the assumption of independent distribution of genes in the parents is most critical to proper interpretation but seems to be least acceptable in actual practice, and (2) that the second assumption of no epistasis, may frequently be incorrect. However, with the proposed diallel analysis involving doubled haploid lines, the epistatic effect is taken into consideration and both additive and additive × additive genetic variances, which are of most concern to plant breeders in self-pollinating crops, can be estimated and the assumption that there is independent gene distribution can be tested. Further, an indication of whether or not the loci are tightly linked is also provided by this analysis. Therefore, using doubled haploid lines to estimate the genetic variances in self-pollinating crops appears to give us better estimates of genetic variances and understanding about the parents and the inheritance of quantitative traits than using F_1 or F_2 bulks. One of the questions that has to be answered before this analytical method can be confidently used is: how many doubled haploid lines should be derived from each cross to make the sampling variance negligible? Reinbergs et al. (1976) showed that at least 20 doubled haploid lines are required to distinguish the yield potential of barley crosses. Computer simulation studies should provide useful guidelines on the optimum sample size for a diallel analysis study.

Acknowledgement

The authors wish to acknowledge that the study was inspired by Dr. R.J. Baker. Thanks are extended to Drs. R.J. Baker, K.J. Kasha, and B.R. Christie for their critical review of this manuscript. Financial support from the National Research Council of Canada is greatly appreciated.

Literature

Baker, R.J.: Issues in diallel analysis. Crop Sci. **18**, 533-536 (1978)

Brim, C.A.: A modified pedigree method of selection in soybeans. Crop Sci. **6**, 220 (1966)

Choo, T.M.; Christie, B.R.; Reinbergs, E.: Doubled haploids for estimating genetic variances and a scheme for population improvement in self-pollinating crops. Theor. Appl. Genet. **6**, 267-271 (1979)

Coughtrey, A.; Mather, K.: Interaction and gene association and dispersion in diallel crosses where gene frequencies are unequal. Heredity **25**, 79-88 (1970)

Jensen, C.J.: Barley monoploids and doubled monoploids: Techniques and experience. In: Barley Genetics III (ed. Gaul, H.) pp. 316-345. München: K. Thiemig 1975

Kasha, K.J.; Kao, K.N.: High frequency haploid production in barley. Nature **225**, 874-876 (1970)

Kasha, K.J.; Reinbergs, E.: Utilization of haploidy in barley. In: Barley Genetics III (ed. Gaul, H.). München: K. Thiemig 1975

Mather, K.; Jinks, J.L.: Biometrical genetics. Ithaca, N.Y.: Cornell Univ. Press 1971

Park, S.J.; Walsh, E.J.; Reinbergs, E.; Song, L.S.P.; Kasha, K.J.: Field performance of doubled haploid barley lines in comparison with lines developed by the pedigree and single seed descent methods. Can. J. Plant Sci. **56**, 467-474 (1976)

Reinbergs, E.; Park, S.J.; Song, L.S.P.: Early identification of superior barley crosses by doubled haploid technique. Z. Pflanzenzüchtg. **76**, 215-225 (1976)

Song, L.S.P.; Park, S.J.; Reinbergs, E.; Choo, T.M.; Kasha, K.J.: Doubled haploid vs the bulk plot method for production of homozygous lines in barley. Z. Pflanzenzüchtg. **81**, 271-280 (1978)

Subrahmanyam, N.C.; Kasha, K.J.: Chromosome doubling of barley haploids by nitrous oxide and calchicine treatments. Can. J. Genet. Cytol. **17**, 573-583 (1975)

Received March 27, 1979
Communicated by R.W. Allard

Dr. E. Reinbergs
Department of Crop Science
University of Guelph
Guelph, Ontario N1G 2W1 (Canada)

Dr. T.M. Choo
Research Station, Agriculture Canada
P.O. Box 1210
Charlottetown
Prince Edward Island C1A 7M8 (Canada)