Theor Appl Genet (1991) 82:54–56

A note on the application of diallel crosses for the analysis of genetic variation in natural populations

M. D. Gebhardt

Zoological Institute, University of Basel, Rheinsprung 9, CH-4051 Basel, Switzerland

Received May 4, 1990; Accepted December 5, 1990
Communicated by J. S. F. Barker

Summary. The complete diallel cross among homozygous lines can be a useful tool to analyze the genetic architecture of natural populations. However, it represents the natural population only approximately, in particular if the number of lines is small and the analyzed traits exhibit inbreeding depression or other forms of directional dominance. Some incorrect expected mean squares that can be found in the literature suggest tests for genetic variance components that can be misleading under such circumstances. Expected mean squares for a factorial analysis and for a modified Hayman analysis are presented and the effect of the number of lines and directional dominance is discussed.

Key words: Quantitative genetics – Diallel cross – Factorial analysis – Hayman analysis – Natural populations

Introduction

Diallel crossing schemes were introduced several decades ago for the genetic analysis of lines used for breeding purposes. More recently, they have also been used to analyze genetic variation in natural populations, because they provide much information on different sources of variation due to the large number of possible comparisons among F_1-families (among lines, half sibs, full sibs, lines used either as paternal or maternal parent, etc.). This is especially true if the full diallel cross is used, which consists of the p^2 possible matings among a set of p parental lines. This note deals exclusively with the full diallel cross of inbred lines and some analyses that have been described (Wearden 1964; Walters and Gale 1977; Mather and Jinks 1982) and applied to the analysis of natural populations (Crusio et al. 1984).

The advantage of so much information has a cost: the number of families to be raised in an experiment increases with the square of the number of parental lines, with the result that this number must often be small for practical reasons. Two problems can then arise: first, the sample of independent genotypes may not be representative of the natural population. Second, and this is the subject of the present note, statistical tests used in some published analyses can be misleading with few parental lines if inbreeding depression is important for the traits of interest.

Current address: Department of Research, Human Genetics, University Hospitals, CH-4031 Basel, Switzerland

Models

From a theoretical point of view, the simplest situation is when the lines are completely homozygous. If the p lines are representative for the gene frequencies in the population, the diallel table can be regarded as a table of gametic combinations, and it can be shown (disregarding environmental variance and reciprocal effects for the moment) that the total variance among family means equals V_G (the genetic variance in the population), the variance among parental array means equals $V_A/2$ (half the additive genetic variance in the population), and the interaction variance equals V_D (the dominance variance) (Mather and Jinks 1982).

A straightforward model reflecting this (allowing also for maternal effects) uses combining abilities of the parental lines (Wearden 1964):

$$y_{ij} = \mu + g_i + g_j + s_{ij} + m_i ,$$

where

y_{ij} = expectation of a cross between inbred line i as mother and line j as father;

μ = the mean of the population;

g_i = the genetic effect of the i^{th} line (general combining ability);

s_{ij} = the interaction between the genetic contributions of parents i and j (specific combining ability);

m_i = the maternal contribution of the i^{th} line.

In the analysis of natural populations, all effects in this model, apart from the mean, are random, with variances σ_g^2, σ_s^2, and σ_m^2. It can be shown that $2\sigma_g^2 = V_A$ and $\sigma_s^2 = V_D$ (Griffing 1956). Wearden (1964) summarizes several alternative ways to analyze the diallel by giving the expectations of a number of sums of squares in terms of the causal variance components defined in this model. Among the analyses that can be constructed from these sums of squares, the factorial and the Hayman analyses are of interest here. For the factorial analysis, where dams and sires are treated as main factors, the following expectations are given for the mean squares (for a single replicate in an experiment with error variance σ^2):

MS (sires): $\sigma^2 + \sigma_s^2 + p\,\sigma_g^2$,

MS (dams × sires): $\sigma^2 + \sigma_s^2$,

corresponding to the relationships among cells in a table of gametic combinations mentioned above. Accordingly, the ratio of these two mean squares would test for additive genetic variance in the population, analogous to the usual factorial design with interaction in stastistics. However, dams and sires are not independent factors as in the usual situation, for their additive genetic contributions are the same, if they come from the same inbred line, and their genetic interactions are symmetrical, i.e., $s_{ij} = s_{ji}$. This caused inconsistencies in a study of the development of *Drosophila* larvae (M.D. Gebhardt and S. C. Stearns, in preparation) when the mean squares were calculated for both the factorial and the Hayman analysis. The traits used were developmental time and weight at eclosion, both life history traits with low heritability, much dominance variance, and strong inbreeding depression in the homozygous lines. The most notable discrepancy was that the test for additive variance was much less frequently significant in the factorial analysis than would be expected from the differences among the inbred lines.

Recalculating the expectations for the mean squares in the factorial analysis revealed that the test suggested by Wearden (1964) is only approximately correct, particularly when p is small. The coefficient of σ_s^2 in the expectation for the correction term in the sums of squares (item I in Wearden's Table 1) seems to have been calculated without considering the symmetry of the genetic interactions. Table 1 shows the correct expectations in the column for the coefficients of σ_s^2, the variance due to dominance effects in this model. It can be seen that the MS(dams × sires) contains a greater contribution of dominance effects by an additional $\sigma_s^2/(p-1)$ than MS(sires). If dominance effects are large in comparison to additive effects and the number of lines is not large, it can make the test overly conservative. The effect of the number of lines can be analyzed further by calculating the expectations for the mean squares in the Hayman analysis. Usually, the b-item (testing for general dominance effects; Hayman 1954) is subdivided into contributions due to mean directional dominance (b1), dominance due to particular parental lines (b2), and dominance specific to each cross (b3). The sums of squares for these items have been rederived by Walters and Gale (1977), based on a slightly different genetic model:

$y_{ii} = a + 2g_i + m_i$,

$y_{ij} = a + g_i + g_j + d_{ij} + m_i$ (for i and j different),

Table 1. Coefficients of causal variance components in two different genetic models for the expectations of mean squares in the factorial and Hayman analysis of the full diallel table with p parents. σ_g^2, σ_s^2, and σ_m^2 are components corresponding to parameters contained in the model given in Wearden (1964). σ_g^2, l^2, σ_{li}^2, σ_{lij}^2, and σ_m^2 correspond to parameters in the model used by Walters and Gale (1977) (definition of these models in the text)

MS	*df*	σ_g^2	σ_s^2	l^2	σ_{li}^2	σ_{lij}^2	σ_m^2
Factorial analysis:							
Dams	$p-1$	p	$(p-1)/p$	0	$(p-2)^2/p$	$(p-2)/p$	p
Sires	$p-1$	p	$(p-1)/p$	0	$(p-2)^2/p$	$(p-2)/p$	0
Dams × sires	$(p-1)^2$	0	$(p-1)/p+1/(p-1)$	$1/(p-1)$	$4/p$	$(p-1)/p+1/p(p-1)$	0
Hayman analysis:							
a	$p-1$	$2p$	$2(p-1)/p$	0	$2(p-2)^2/p$	$2(p-2)/p$	$p/2$
b	$p(p-1)/2$	0	$2(p-1)/p+2/p^2$	$2/p$	$8(p-1)/p^2$	$2(p-2)/p+4/p^2$	0
b1	1	0	$(p+1)/p$	$p-1$	$4(p-1)/p$	$2/p$	0
b2	$p-1$	0	$(p+2)/p$	0	$4(p-2)/p$	$4/p$	0
b3	$p(p-3)/2$	0	2	0	0	2	0

where
a = the mean of the inbred lines;
g_i = the genetic effect of the i^{th} line;
d_{ij} = the dominance deviation;
m_i = the maternal contribution of the i^{th} line.

Dominance deviations are measured from the mid-parent value in this model and do not represent specific combining abilities. They can be subdivided as

$$d_{ij} = l + l_i + l_j + l_{ij},$$

where l, l_i, l_{ij} correspond to the b1, b2, and b3 items in the Hayman analysis. Here, the dominance components l_i and l_{ij} are assumed to be random variables in the population, with zero expectation and respective variance σ^2_{li} and σ^2_{lij}. Parameter l estimates a fixed effect and contributes to the variation among F_1 families by its square (Table 1). The expectations for the mean squares in the factorial and Hayman analyses under this model are given in Table 1. The a and b items in the Hayman analysis are identical to Griffing's (1956) general and specific combining abilities analyzed with Method 1 under Model II.

Comparing the expectations for the mean squares under the two models leads to the following conclusions. Both MS(dams × sires) and MS(b) test for dominance effects in general. However, inbreeding depression (mean directional dominance, parameter l) will inflate the mean squares if p is small. This is also true for the dominance component that is due to particular lines (parameter l_i). If the number of lines is large, these contributions will be trivial and both mean squares (dams × sires and b) will test for the dominance variance in the population, V_D. If the traits in the analysis do not exhibit directional dominance, then all subdivisions of b (b1, b2, and b3) test for V_D in the population (seen in the column for σ^2_s). If there is directional dominance, only the b3-item yields a valid test for V_D. This is because dominance variation in an outcrossing natural population is not a function of inbreeding depression (the difference between the mean performances of homozygous and heterozygous individuals), but of the specific interaction between different alleles at each locus. There are generally no completely homozygous individuals in a randomly breeding population.

Another observation can be made if the subdivisions of b are compared. If one is interested in effects of inbreeding depression, then the b1-item is normally used to test for the difference between homozygous and heterozygous individuals. However, Table 1 shows that the mean square of b1 is inflated by variation in directional dominance among lines (l_i) as well as by (for small p) specific dominance effects (l_{ij}). Only approximate tests for mean directional dominance are possible with the items in the Hayman analysis: MS(b1) over MS(b2) for large p, or the test of MS(b1) over an appropriate error term if MS(b2) and MS(b3) are not significant.

Conclusion

Most of the above complications can be avoided if the selfs are not included in a diallel analysis of a natural population. This was also stated by Griffing (1956), who did not elaborate on the reasons for this, however. The progeny produced by a complete diallel scheme represent the original population accurately only if directional dominance is not important. In an outcrossing natural population, directional dominance is not important for quantitative traits because individuals homozygous at all loci contributing to the trait do not occur. In contrast, the proportion of homozygous families is considerable in a complete diallel scheme, and the contribution of the difference between the homozygous and the heterozygous families to the total variance may be substantial. The full diallel cross has been used in studies of genetic variation in natural populations, with the Hayman analysis often being applied (which allows inference on V_D in the population if only b3 is used as a criterion), but also a factorial analysis as suggested by Wearden (1964). Depending on the nature of the traits (whether inbreeding depression is important or not) and the number of lines used, this can lead to erroneous conclusions regarding the relative importance of genetic variance components.

Acknowledgements. I thank S. Henrich, A. van Noordwijk, S. C. Stearns, and an anonymous reviewer for comments on earlier drafts of the manuscript. Financial support was provided by the Stiftung Emilia Guggenheim-Schnurr der Naturforschenden Gesellschaft in Basel and the Werenfels-Fonds (Freiwillige Akademische Gesellschaft, Basel).

References

Crusio WE, Kerbusch, JML, Abeelen JHF van (1984) The replicated diallel cross: a generalized method of analysis. Behav Genet 14:81–104

Griffing B (1956) A generalized treatment of the use of diallel crosses in quantitative inheritance. Heredity 10:31–50

Hayman, BI (1954) The analysis of variance of diallel tables. Biometrics 10:235–244

Mather K, Jinks JL (1982) Biometrical genetics, 3rd edn. Chapman and Hall, London

Walters DE, Gale JS (1977) A note on the Hayman analysis of variance for a full diallel table. Heredity 39:401–407

Wearden S (1964) Alternative analyses of the diallel cross. Heredity 19:669–680