A GENERAL METHOD OF DETECTING ADDITIVE, DOMINANCE AND EPISTATIC VARIATION FOR METRICAL TRAITS

I. THEORY

M. J. KEARSEY and J. L. JINKS Department of Genetics, The University, Birmingham, 15

Received 2.xii.67

1. INTRODUCTION

In any attempt to obtain efficient estimates of the additive, dominance and environmental components of variation for a metrical trait from second degree statistics three serious difficulties inevitably arise. Firstly it is assumed in most analyses that non-allelic interactions are absent although these analyses rarely provide a valid test of this assumption. There are obvious exceptions to this. Among the multiple mating designs the sole exception is the diallel analysis of Jinks and Hayman (Jinks, 1956; Hayman, 1954) in which both the consistency of $W_r - V_r$ over arrays and information derived from F₂ and back-cross generations provide tests of the adequacy of the additive, dominance model. Another exception applies only to crosses between single pairs of inbred lines and the generations derived from them by selfing, back-crossing, sib-mating, etc. (Mather and Vines, 1952; Opsahl, 1956). To determine the adequacy of the simple model at the level of second-degree statistics, in such a situation, necessitates least squares estimates of components from variances derived from different generations with all the inherent difficulties of significance tests that such comparisons entail.

Secondly, estimates of dominance components invariably have much larger standard errors than do the corresponding additive components. Thirdly, these additive and dominance components are differentially affected by linkage and correlated gene distributions in the parents, and are only comparable in the unlikely event of the population sample being in linkage equilibrium.

An experimental method designed to overcome the second point above was described by Comstock and Robinson (1952) in their Experiment III, and permitted an analysis of generations derived by randomly mating an F_2 from two inbred lines. The purpose of this paper is to describe a simple extension of this design so as to provide not only the more efficient estimates of dominance, but also an unambiguous test for epistasis. The use of this extended design to investigate populations other than F_2 's and their randomly mated derivatives will be discussed. Indeed it will be shown that this approach has a general validity for investigating any population irrespective of both gene frequencies or mating system.

2. Model

All gene effects will be described by the $F \propto$ metric discussed by Van der Veen (1959). In this, 'd' and 'h' have the same meaning as the 'u' and 'au' of Comstock and Robinson, while 'i', 'j', and 'l' are the digenic interaction terms ascribable to homozygous × homozygous, homozygous × heterozygous, and heterozygous × heterozygous pairs of loci. These gene effects are, in this case, defined as deviations from the mean of all possible inbred lines that could be obtained from the lines used.

3. Experimental design

A random sample of n individuals from the population to be investigated are crossed, as male parents to the same three testers. Of these testers, two (L₁ and L₂) are inbred lines as in the conventional Experiment III, while L₃ is the F₁ produced from them. The experiment will then consist of 3n families, each family being replicated by raising either r plots or rindividuals in a randomised block design. In the case of self-fertile species reciprocal crosses may be made, but they will not be considered further here.

4. ANALYSIS

It is convenient to divide the analysis into two parts, the one concerned with the test for epistasis and the other concerned with testing for and estimating genetic components.

Let us assume k loci, segregating in the population, and of these k' (where $k' \leq k$) differ between L_1 and L_2 . If k' < k, the test for epistasis will not be invalidated and will at least detect epistasis between these k' loci, and the same is true for dominance. If one assumes that these k' loci are a random sample of the loci segregating in the population, then this analysis will indicate whether or not epistasis and dominance are components of the genetic architecture but will not indicate their absolute contribution. However, in order that the estimate of additive variation will have any real meaning it is essential that k' < k (*i.e.* same loci).

(i) Test for epistasis

To explain this test, we shall consider the situation k' = 2. Table 1 shows the expected mean values of families derived by crossing each of the 9 possible genotypes (for 2 loci) present in the population with the three testers. To each of the means in table 1 (a) should be added a term c_i which is constant for any one row, but will differ between rows, and accounts for the additive and dominance contributions of the k - k' loci which L_1 and L_2 have in common.

If for the *i*th individual sampled from the population one computes $L_{1i}+L_{2i}-2L_{3i}$ (table 1 (b)), the c_i , d, and h terms cancel and epistatic terms alone remain (where L_{1i} refers to the mean of the cross $i \neq L_1$, etc.). Furthermore, this is true for any number of loci and is not confined to digenic interactions. This must be so since for any one row $L_1+L_2 = 2L_3$ for the c_i , d_i and h_i terms and is therefore independent of the degree of inbreeding, gene frequencies, gene correlation, etc. Thus, for any base population $L_{1i}+L_{2i}-2L_{3i}$ should not be significantly different from zero, in the absence

of epistasis, for every set of crosses. Although the above test was developed with L_1+L_2 in association, the same result is obtained for dispersion.

Thus the average variance of $(L_{1i}+L_{2i}-2L_{3i})$ should not be significantly greater than the replicate error if epistasis is absent. However, as will be seen later, the replicate error of L_{3i} may well differ from that of L_{1i} and L_{2i} for genetical reasons and so in general the correct test of significance is

$$\frac{r\sum_{i=1}^{n} (L_{1i} + L_{2i} - 2L_{3i})^2}{n(VL_1 + VL_2 + 4VL_3)} \text{ for } n, \text{ and } 3n(r-1) \text{ d.f.}$$

where VL_1 = average replicate variance of L_1 , etc.

Only if epistasis is found not to be significant is it useful to attempt to estimate the additive and dominance components as shown below.

TABLE 1

The contribution of main effects and digenic interaction parameters to the means of L_{1i} , L_{2i} , L_{3i} and $L_{1i} + L_{2i} - 2L_{3i}$ for each of the 9 possible genotypes in respect of genes A-a and B-b.

Freq. $adb ha hb i j l da db ha hb i j l da db ha hb i j l da db ha hb i f_{11} AABB 1 1 1 1 1 1 \frac{1}{2} \frac{1}{2$	<i>(a)</i>	$L_1(AABB)$ $L_2(aabb)$		$L_3(AaBb)$	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Freq.	da db ha hb i j l	da db ha hb i j l	da db ha hb i j	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	f ₁₁ AABB f ₁₂ AABb f ₁₃ AAbb f ₂₁ AaBB f ₂₂ AaBb f ₃₃ Aabb f ₃₁ aaBB f ₃₂ aaBb f ₃₃ aabb	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	

<i>(b)</i>	L_1	$L_1 + L_2 - 2L_3$		
Freq.	i	j	l	
f ₁₁ AABB	$\frac{1}{2}$	-1	1/2	
f12 AABb	1	$-\frac{1}{2}$		
f ₁₃ AAbb	12			
f21AaBB	<u> </u>			
f22 AaBb	호	-	_	
f ₂₃ Aabb	12	1/2	-	
f _{a1} aaBB	12	_	$-\frac{1}{2}$	
f_{32} aaBb	12	1	_	
f ₃₃ aabb	12	ī	$\frac{1}{2}$	

(ii) Additive and dominance components

Dominance, as measured in this design, refers only to the k' loci for which L_1 and L_2 differ. Thus if one is only interested in finding whether dominance is present or absent in a population, virtually any pair of inbred lines will suffice. If on the other hand, one wishes to obtain an estimate of the total dominance variation, then L_1 and L_2 must differ at all the loci which are segregating in the population. With the exception of an F_2 or its derivatives, therefore, they would have to be high and low selections from that population. Since the F_2 situation has been discussed by Comstock and Robinson (1952) we shall confine our attention to other situations.

The analysis of additive and dominance effects is exactly as described by Comstock and Robinson, *i.e.* by obtaining the variances of sums and differences of L_{1i} and L_{2i} , except that σ_{m1}^2 has been redefined and is equal to $\frac{1}{2}$ σ_{m1}^2 of Comstock and Robinson (table 2.) This change was made simply to facilitate comparisons between additive and dominance components.

TABLE 2

Analysis of variance

Item	d.f.	e.m.s.
Sums $(L_1 + L_2)$	n-1	$\sigma^2 + 2r\sigma_m^2$
Diffs $(L_1 - L_2)$	n-1	$\sigma^2 + 2r\sigma_{m1}^2$
Within crosses	n(r-1)	σ^2

The interpretation of these $\sigma^{2^{\circ}s}$ are as shown in table 3 for the various types of base population to be considered below.

- (a) A population in which nothing is known of the mating system, gene frequencies or gene correlations. Here α , β , γ are the frequencies of AA, Aa, aa zygotes respectively and t_{jk} and t'_{jk} are the coefficients of covariance terms which are functions of the degree of linkage equilibrium between the *j*th and *k*th loci.
- (b) A population produced by at least one generation of random mating, with gene frequencies and correlation unknown. Here D_{jk} is the linkage disequilibrium between the *j*th and *k*th loci in the gametes producing this population, and *u*, *v* are the allele frequencies.
- (c) A set of inbred lines with unknown gene frequencies and correlations, again with D_{jk} denoting the linkage disequilibrium.

The special case of equal gene frequencies can easily be derived from table 3 by putting $u = (\alpha + \frac{1}{2}\beta) = v = (\gamma + \frac{1}{2}\beta) = \frac{1}{2}$.

It can be seen from table 3 that σ_{m1}^2 estimates dominance irrespective of the type of pupulation involved, but its efficiency will depend on the gene distribution in L_1+L_2 and in the population. One can imagine various combinations of gene correlations in the parents (\pm) and D_{jk} , and it is obvious that the absolute magnitude of σ_{m1}^2 is as much affected by these as by dominance. All one can say with certainty, and this is true of all other designs, is that if σ_{m1}^2 is significant then there is dominance at least at some loci. Non-significance of σ_{m1}^2 on the other hand does not necessarily indicate no dominance, although this situation is unlikely to arise providing L_1 and L_2 are relatively distinct phenotypically.

If k' = k, which in terms of unknown populations infers that L_1 and L_2 are extreme high and low selection lines, the interpretation can be taken further. Selection lines must have a high degree of gene association, *i.e.* one can replace the $\pm \Sigma h_j h_k$ with $+\Sigma h_j h_k$, which in all the cases considered makes the coefficients within σ_{m1}^2 identical with σ_m^2 . Thus although both σ_m^2 and σ_{m1}^2 are inflated by excess coupling linkages in L_1 and L_2 they are affected to the same extent, and the ratio, σ_{m1}^2/σ_m^2 , is a measure of average dominance irrespective of gene distributions in the population.

TABLE 3

The expectations of σ_m^2 , σ_{m1}^2 and σ^2 for various types of population (see text for description of (a), (b) and (c)).

$$(a) \quad \sigma_m^2 = \frac{1}{2} \sum_{j} [(\alpha + \gamma)\beta + 4\alpha\gamma] d_j^2 + \sum_{jk} t_{jk} d_j d_k$$

$$\sigma_{m1}^2 = \frac{1}{2} \sum_{j} [(\alpha + \gamma)\beta + 4\alpha\gamma] h_j^2 \pm \frac{1}{2k} t_{jk} h_j h_k$$

$$\sigma^2 = \frac{1}{2} \sum_{j} \beta_j d_j^2 + \frac{1}{2k} \sum_{j} \beta_j h_j^2 + \sum_{jk} t'_{jk} (1 - 2r^*) (h_j h_k \pm d_j d_k)$$

$$(b) \quad \sigma_{m}^{2} = \frac{1}{2} \sum_{j} uvd_{j}^{2} + \sum_{jk} D_{jk}d_{j}d_{k}$$

$$\sigma_{m1}^{2} = \frac{1}{2} \sum_{j} uvh_{j}^{2} \pm \sum_{jk} D_{jk}h_{j}h_{k}$$

$$\sigma^{2} = \frac{1}{2} \sum_{j} uvd_{j}^{2} + \frac{1}{2} \sum_{j} uvh_{j}^{2} + \sum_{jk} (1 - 2r^{*})D_{jk}(h_{j}h_{k} \pm d_{j}d_{k})$$

(c)
$$\sigma_m^2 = \sum_j uvd_j^2 + 2\sum_{jk} D_{jk}d_jd_k$$
$$\sigma_{m1}^2 = \sum_j uvh_j^2 \pm 2\sum_{jk} D_{jk}h_jh_k$$
$$\sigma^2 = 0$$

* r = recombination fraction. † \pm indicates coupling (+) or repulsion (-) in L₁+L₂.

5. DISCUSSION

It has been shown that this design has several advantages over other multiple mating designs. It provides a test for epistasis and dominance, and, as long as L_1 and L_2 are extreme selections from the population, it provides estimates of additive and dominance variation with equal precision. Since each male parent is crossed to three testers only, it allows one to investigate a larger sample from the population, for the same experimental effort, than does any other design. The three testers are inbred lines or F_1 's and hence can be replicated many times, so avoiding the necessity of multiple matings to the same female.

However it is necessary to consider how generally this design can be applied. Since estimates of additive variation have meaning only if L₁ and L₂ are extreme selection lines, heritability estimates can be obtained only after selection has taken place. Thus is is of little use to breeders wishing to predict a selection response. It is felt that the principal use of this design is as a tool for investigating the genetic architecture of traits in natural populations. Here one is not so much interested in the total genetic variation of a given character in a population, as in the type of gene action or interaction that is present, and attempting to relate this to natural selection and ecology. Providing epistasis and/or dominance is present, it is axiomatic that there is additive variation. Any pair of lines, preferably phenotypically divergent, and from the same population, will detect these sources of variation for the sample of loci for which the testers differ. If it is feared that this pair of lines might be a biased sample the experiment can be repeated using different testers and male parents. In the event of nonadditive variation, as measured by this design, being absent, a more detailed

examination of the additive variation can be undertaken with any of the many other designs appropriate to the material.

It is necessary to stress that the tester lines should be derived from the population to be studied. One aim, at least, of studies of genetic architecture, is to relate the architecture to selection, and implicit in this is the idea that gene action and interaction are affected by selection to some extent. Thus the models developed above are valid only within a population and cannot cope with, for example, the same allele showing different dominance properties in different populations. Such a situation would almost certainly be detected as epistasis, but it would not in fact be a type of epistasis found within a population. It is a well-known phenomenon that wide crosses exhibit a type of non-additive variation not present within a population and this may well be a manifestation of this problem.

If the population consists of inbred lines, which have to be raised continually to maintain the material, then the line values themselves can be substituted for the crosses to L_3 to provide a test for epistasis. If P_i is the mean of the *i*th inbred line, then:

$$L_{1i}+L_{2i}-P_i = \mu + \Sigma h + f_i(i) + g_i(l) = (dev)_i$$

where f_i and g_i are coefficients which depend on the genotype of the lines. Since here the expected value of $(\text{dev})_i = \mu + \Sigma h$ (and not 0) in the absence of non-allelic interaction, the appropriate variance is computed around the mean of all deviations and has n-1 d.f. (c.f. test for epistasis above).

If all the generation considered above, *i.e.* crosses to L_1 , L_2 , L_3 , are selfed, the tests for epistasis, dominance and additive effects still hold, and the coefficients of the dominance and epistatic terms alone change as simple algebra will show. This fact could well be made use of in material which naturally selfs but which produces little F_1 seed for technical reasons (*e.g.* wheat).

6. Summary

1. A method is described of detecting additive dominance, and epistatic variation in a population, by crossing males from this population to two inbred lines (L_1+L_2) and their F_1 , (L_3) .

2. Irrespective of the genetic constitution of this population (*i.e.* gene frequencies, linkage disequilibrium, etc.), the method will detect dominance and epistasis for those loci for which L_1 and L_2 differ.

3. Provided that no epistastis is detected and L_1 and L_2 are extreme high and low selection lines for the character investigated, the method also allows one to estimate additive and dominance components with equal precision and obtain estimates of average dominance.

4. The analyses of various types of population and their genetic interpretation are described.

5. Variations of this method especially appropriate to self-fertile and naturally inbreeding material are discussed.

6. It is considered that this crossing scheme and its analysis may well be a very useful method of investigating the genetic architecture of natural populations.

Acknowledgment .- This work was supported by the Agricultural Research Council.

7. References

- COMSTOCK, R. E., AND ROBINSON, H. F. 1952. Estimation of average dominance of genes. Heterosis, Chap. 30. Iowa State College Press.
- HAYMAN, B. I. 1954. The theory and analysis of diallel crosses. Genetics, 39, 789-809.
- JINKS, J. L. 1956. The F₂ and backcross generations from a set of diallel crosses. *Heredity*, 10, 1-30.
- MATHER, K., AND VINES, A. 1952. The inheritance of height and flowering time in a cross of Nicotiana rustica. Quantitative Inheritance, pp. 45-80. H.M.S.O.
- OPSAHL, B. 1956. The discrimination of interactions and linkage in continuous variation. Biometrics, 12, 415-432.
- VAN DER VEEN, J. H. 1959. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. *Genetica*, 30, 201-232.