

THE EFFICIENCY OF THE NORTH CAROLINA EXPT. III AND THE SELFING, BACKCROSSING SERIES FOR ESTIMATING ADDITIVE AND DOMINANCE VARIATION

M. J. KEARSEY

Department of Genetics, University of Birmingham, Birmingham B15 2TT, England

Received 12.ii.80

SUMMARY

The optimum family structure of two experimental designs ((i) Non segregating generations plus F_2 's and backcrosses; (ii) The North Carolina Expt. III) have been investigated with respect to the estimation of additive (D) and non additive (H) genetic variation for a quantitative trait. In the former design it is shown that the generations should ideally be replicated in proportion to the total variation of each generation (if this is known) while in the latter, one should aim to sample at least 20 F_2 plants. The relative efficiencies of the two designs are compared and the considerable advantages of the N.C. Expt. III design are illustrated and emphasised.

1. INTRODUCTION

AN important problem in initiating a breeding programme starting from a collection of inbred lines is to form an objective judgement of which crosses are likely to produce the most promising inbred lines on inbreeding the F_2 . Such judgements are based on information about the mean and genetic variance of the cross, the latter involving either the additive genetic variance alone (D) or the inverse dominance ratio ($\sqrt{D/H}$) (Jinks and Perkins, 1972). The present paper is concerned with a consideration of how to estimate these effects (D and H) with the greatest precision; D and H being as defined by Mather and Jinks (1971).

Since we are dealing with generations derived from two inbred lines there are basically two experimental designs which are appropriate—the Expt. III of Comstock and Robinson (1952) which will be referred to as the North Carolina Expt. III and secondly F_2 and backcross generations (Mather and Jinks, 1971). Not only are they relatively simple to execute but they provide information on D , H and E as well as on epistasis and genotype-environment interaction ($G \times E$). As their power in testing the adequacy of the simple additive dominance model is discussed elsewhere (Kearsey and Jinks, 1968; Jinks and Perkins, 1969; Kearsey, 1970; Pooni and Jinks, 1976), it will be assumed that epistasis and $G \times E$ are absent and we will concentrate on the estimation of D and H .

With respect to these two types of experiment we may ask the following questions.

(1) For a given genetic model (narrow heritability and dominance ratio) and a fixed amount of effort, what is the best distribution of resources into different families in order to estimate D and H with the minimum variance?

(2) Does the optimal structure vary with the genetical control and the parameters (D and H) to be estimated? If so, is there a realistic compromise structure that might be used in the absence of any prior information?

(3) What are the relative advantages of the 2 types of experiment?

(4) How large an experiment is it necessary to raise in order to achieve a given level of estimate reliability?

We will now consider each of these questions in turn.

2. OPTIMAL STRUCTURES

(i) F_2 's and backcrosses

If we raise F_1 's and F_2 's and backcrosses it is possible to obtain variances which have the expectations shown in table 1(a).

TABLE 1

*The statistics and their genetical expectations from the two experimental designs.
(a) F_2 and backcross experiment, (b) North Carolina Experiment III*

(a)

Generation	Observed variance	Degrees of freedom n_i	Components of variation			
			D	H	F	E_1
Non-segregating (P_1, P_2 or F_1)	V_E	n_1	0	0	0	1
F_2	V_{F_2}	n_2	0.5	0.25	0	1
B_1	V_{B_1}	n_3	0.25	0.25	-0.5	1
B_2	V_{B_2}	n_4	0.25	0.25	0.5	1

(b)

ANOVA

Item	df	MS	e.m.s.
Additive	$k-1$	MS_A	$\sigma^2 + 2r\sigma_A^2$
Non additive	$k-1$	MS_D	$\sigma^2 + 2r\sigma_D^2$
Replicate error	$2k(r-1)$	MS_R	σ^2
Where $\sigma_A^2 = \frac{1}{8}D$			
$\sigma_A^2 = \frac{1}{8}H$			
$\sigma^2 = \frac{1}{8}D + \frac{1}{8}H + E_1$			

(a) The additive variation (D) can be estimated from the F_2 's and backcrosses alone, *viz.*,

$$\hat{D} = 2[2V_{F_2} - V_{B_1} - V_{B_2}] \quad (1)$$

while the expected variance of V_D is:

$$V_{\hat{D}} = 8 \left[\frac{4V_{F_2}^2}{n_2} + \frac{V_{B_1}^2}{n_3} + \frac{V_{B_2}^2}{n_4} \right] \quad (2)$$

where n_i are the degrees of freedom of each generation. If N is the total degrees of freedom ($= n_2 + n_3 + n_4$) then we can rewrite (2) in terms of p_i , the

proportion of the total degrees of freedom associated with the i th generation. Thus

$$V_D = \frac{8}{N} \left[\frac{4V_{F_2}^2}{p_2} + \frac{V_{B_1}^2}{p_3} + \frac{V_{B_2}^2}{p_4} \right] \\ = \frac{8}{N} [Q].$$

Such that $p_2 + p_3 + p_4 = 1.0$.

We can arrive at those values of p_i which minimise V_D by differentiating Q with respect to p_2 and p_3 . This yields the result:

$$\left. \begin{aligned} p_2 &= \frac{2V_{F_2}}{2V_{F_2} + V_{B_1} + V_{B_2}} \\ p_3 &= \frac{V_{B_1}}{2V_{F_2} + V_{B_1} + V_{B_2}} \\ p_4 &= \frac{V_{B_2}}{2V_{F_2} + V_{B_1} + V_{B_2}} \end{aligned} \right\}. \quad (3)$$

Under such conditions substituting equations (3) into (2) yields

$$V_D = \frac{8}{N} [2V_{F_2} + V_{B_1} + V_{B_2}]^2. \quad (4)$$

We can now estimate the optimal proportions of the three different generations and V_D for various genetical situations from equations (3) and (4) and these are shown in table 2(a). In this table we have included all combinations of four values of the narrow heritability, three values of the dominance ratio

TABLE 2

(a) *Optimum proportions of F_2 and backcross individuals for estimating D under various genetical models. The variance of D (V_D) is given at the bottom of the table for these optimal designs and is expressed as a function of $N/V_{F_2}^2$ where N and V_{F_2} are respectively the total degrees of freedom and the phenotypic variance of the F_2* (b) *Minimum experimental size to achieve $D/\sqrt{V_D} = 1.96$*

(a)			h_n^2			
h/d	F/\sqrt{DH}		0.2	0.4	0.6	0.8
0	0	F_2	0.53	0.56	0.59	0.62
0.5	0	B_1	0.24	0.22	0.20	0.19
1.0	0	B_2	0.24	0.22	0.20	0.19
0.5	1.0	F_2	0.53	0.56	0.59	0.62
0.5	1.0	B_1	0.21	0.17	0.12	0.06
0.5	1.0	B_2	0.26	0.28	0.29	0.31
1.0	1.0	F_2	0.53	0.56	0.59	—
1.0	1.0	B_1	0.18	0.11	0.03	—
1.0	1.0	B_2	0.29	0.33	0.38	—
$NV_D/(V_{F_2})^2$			115.52	103.68	92.48	81.92
(b)			N_{\min}			
			2774	622	247	123

with either zero or complete correlation of d and h ; (as measured by $F/\sqrt{D \cdot H}$). Since the actual size of V_D will depend on N and V_{F_2} , V_D is expressed as $NV_D/V_{F_2}^2$.

(b) From a consideration of table 1 it can be seen that in order to estimate H , we need to include the non-segregating generations (the parental inbreds and their F_1) whose combined variance will be referred to as V_E . Thus:

$$\hat{H} = 4(V_{B_2} + V_{B_1} - V_{F_2} - V_E) \quad (5)$$

and

$$V_{\hat{H}} = 32 \left[\frac{V_{B_2}^2}{n_2} + \frac{V_{B_1}^2}{n_3} + \frac{V_{F_2}^2}{n_4} + \frac{V_E^2}{n_1} \right]. \quad (6)$$

Again, assuming that, of the total degrees of freedom, N' ($= n_1 + n_2 + n_3 + n_4$), a proportion p_i is associated with the i th generation, we obtain values of p_i which minimise $V_{\hat{H}}$ as follows:

$$\left. \begin{aligned} p_1 &= V_E/VT \\ p_2 &= V_{F_2}/VT \\ p_3 &= V_{B_1}/VT \\ p_4 &= V_{B_2}/VT \end{aligned} \right\} \quad (7)$$

TABLE 3

(a) Optimum proportions of nonsegregating generations (N.S.), F_2 and backcross individuals for estimating H under various genetical models (cf. table 2) (b) Minimum experimental size to achieve $H/\sqrt{V_{\hat{H}}} = 1.96$

(a)

h/d		$F/\sqrt{D \cdot H}$	h_n^2			
			0.2	0.4	0.6	0.8
0.5	0	N.S.	0.22	0.17	0.12	0.04
		F_2	0.28	0.32	0.37	0.43
		B_1	0.25	0.25	0.26	0.26
		B_2	0.25	0.26	0.26	0.26
		$NV_H/(V_{F_2})^2$	409	318	238	169
0.5	1.0	N.S.	0.22	0.17	0.12	0.04
		F_2	0.28	0.32	0.37	0.43
		B_1	0.22	0.19	0.15	0.09
		B_2	0.28	0.32	0.37	0.43
		$NV_H/(V_{F_2})^2$	409	318	238	169
1.0	0	N.S.	0.20	0.13	0.04	—
		F_2	0.29	0.33	0.40	—
		B_1	0.26	0.27	0.28	—
		B_2	0.26	0.27	0.28	—
		$NV_H/(V_{F_2})^2$	392	288	200	—
1.0	1.0	N.S.	0.20	0.13	0.04	—
		F	0.29	0.33	0.40	—
		B	0.20	0.13	0.04	—
		B	0.31	0.40	0.52	—
		$NV_H/(V_{F_2})^2$	392	288	200	—

(b)

h/d					
0.5	N_{\min}	157114	30495	10143	4064
1.0	N_{\min}	9412	1729	534	—

Where

$$V_T = V_E + V_{B_1} + V_{B_2} + V_{F_2}.$$

Using the optimal proportions, $V_{\hat{H}}$ becomes

$$V_{\hat{H}} = \frac{32}{N'} (V_{B_1} + V_{B_2} + V_{F_2} + V_E)^2. \quad (8)$$

The optimal structures and corresponding variances of H are shown in table 3(a).

(ii) *The North Carolina Expt. III*

In the basic N.C. Expt. III of Comstock and Robinson (1952) each F_2 individual is crossed (normally as the male parent) to both of its inbred grandparents. Thus a typical experiment will consist of 2 families from every one of the k F_2 parents, each family consisting of r replicates. The analysis of these data is as shown in table 1(b).

From the mean squares in the analysis of variance we can estimate D , H and their variances as follows:

$$\hat{D} = 8 \frac{(MS_A - MS_R)}{2r}$$

$$\hat{H} = 8 \frac{(MS_D - MS_R)}{2r}$$

$$V_{\hat{D}} = \frac{32}{r^2} \left(\frac{MS_A^2}{k-1} + \frac{MS_R^2}{2k(r-1)} \right)$$

$$V_{\hat{H}} = \frac{32}{r^2} \left(\frac{MS_D^2}{k-1} + \frac{MS_R^2}{2k(r-1)} \right).$$

Since the family size, r , appears in the expectation of both MS_A and MS_D as well as in the degrees of freedom for MS_R it is not possible to obtain general solutions for r and k which minimise these last two equations. However, if we assume that the total experimental size is sufficiently large, we can substitute k for $(k-1)$ and solve the equations numerically. Using this approach we obtain the values of r necessary to minimise $V_{\hat{D}}$ and $V_{\hat{H}}$ shown in table 4 together with the corresponding values of $V_{\hat{D}}$ and $V_{\hat{H}}$ expressed as a function of $N/V_{F_2}^2$, where V_{F_2} is the phenotypic variance of the F_2 .

Armed with these values we can estimate N for any given degree of precision. For example, if we wish to estimate D and H such that they are both greater than twice their standard errors, then since

$$D = 2h_n^2 V_{F_2}$$

$$H = 2(h/d)^2 h_n^2 V_{F_2}$$

the experimental sizes N_D , N_H , required to achieve the above level of accuracy are

$$N_D = \frac{3.8416NV_D}{4h_n^4 V_{F_2}^2} \quad \text{and} \quad N_H = \frac{3.8416NV_H}{4(h/d)^2 h_n^4 V_{F_2}^2}$$

where 3.8416 is the square of the normal deviate for 5 per cent probability.

TABLE 4

N.C. Experiment III. Optimal replicate sizes (r), parameter variances V (expressed as $NV_D/V_{F_2}^2$, $NV_H/V_{F_2}^2$), minimal experimental sizes (N), and corresponding values of k necessary to achieve significance

Parameter	h/d		h_n^2			
			0.2	0.4	0.6	0.8
D	0	r	9	4	2	2
		V	23	37	48	51
		N	558	448	136	80
		k	31	28	34	20
D	0.5	r	9	4	2	2
		V	23	37	45	47
		N	558	448	120	72
		k	31	28	30	18
H	0.5	r	35	14	7	4
		V	6	9	10	10
		N	2310	868	434	240
		k	33	31	31	30
D, H	1.0	r	8	3	2	—
		V	22	34	34	—
		N	528	204	92	—
		k	33	34	23	—

This approach yields the values of N_D and N_H and hence k shown in table 4, after adjusting to ensure that k and r are integers. It can be seen from this table that the optimum value of k varies over a narrow range (18-34). In fact, if one calculates the minimum experimental size required with different values of k (table 5) one finds that it maintains a relatively constant, low value in the range $k = 20$ to 40 but increases rapidly if k falls

TABLE 5

N.C. Experiment III. Minimal experimental sizes (N) required in order that the estimated component (D or H) should be 1.96 times its standard error for various values of k (the number of F_2 parents) and h_n^2 (the narrow heritability). (a) Estimation of D , (b) estimation of H . ($N = 2kr$)

k	h_n^2			
	0.2	0.4	0.6	0.8
(a)				
10	2080	860	460	260
15	600	330	180	90
20	600	280	160	80
25	600	250	150	100
30	600	240	180	120
35	560	280	140	140
40	560	240	160	160
(b)				
10	7000	3000	1660	820
15	2880	1140	600	270
20	2400	960	480	280
25	2250	900	500	250
30	2160	840	480	240
35	2170	910	490	280
40	2160	880	480	240

below 20. Given this property it is therefore expedient always to aim to sample at least 20 F_2 individuals, irrespective of the genetical properties of the population and the parameter to be estimated, and to adjust the number of replicates in line with the heritability.

3. DEPENDENCE OF STRUCTURES ON GENETICAL CONTROL AND PARAMETER TO BE ESTIMATED

(i) F_2 's and backcrosses

It is clear from table 2 that the relative proportions of F_2 's to backcrosses, appropriate for estimating D , are comparatively constant (48-38 per cent backcross individuals) over a wide range of genetic situations. However the more potence there is, the more individuals of the backcross to the recessive parent should be raised at the expense of the backcross to the dominant parent. A similar situation holds with respect to estimating H (table 3), the proportion of the experiment devoted to backcrosses varying from 50 to 56 per cent with the same dependence on potence as before. Of the remaining individuals the proportion from the non-segregating generations declines sharply with an increase in the heritability.

It is interesting to note that the optimum strategy for estimating dominance is precisely the strategy proposed by Jinks and Perkins (1969) for use with first degree statistics, *i.e.*, to adjust the number of individuals in each generation so as to make the variance of every generation mean the same.

A comparison of table 2 and 3 shows that the optimum strategies for estimating D and H are different; for " D " we require more F_2 's than backcrosses while for " H " the reverse is true, with almost twice as many backcross individuals being required when heritabilities are high. Given that we wish to estimate both D and H from the same experiment, is there a convenient compromise in the design?

In all examples illustrated in table 2(a) and 3(a), the variance of our estimate of " H " for a given sized experiment is at least twice and in some cases up to 4 times greater than the corresponding value for " D ". This is in line with intuition since a breeding system involving selfing, in that it reduces heterozygosity, is not ideal for the study of dominance. It does suggest that one might afford to lose a little sensitivity in estimating D in order to obtain more reliable values for H . In fact if we adopt a design incorporating the proportions shown in table 3 for optimising dominance estimation we find that our variances for D increase but only by 12 to 14 per cent over the range of genetical models examined. This would seem to be further justification for adopting the approach of Jinks and Perkins (1969), although it does assume some prior knowledge of the likely variances of the various generations.

In the absence of any firm knowledge of the likely genetical control of the characters, we can do no better than adopt the proportions suggested for an intermediate level of heritability and dominance (e.g., 0.4, 0.5 respectively). It would appear that unless either the heritability or potence are very high there is less than a 20 per cent increase in variance with such a design; indeed in most cases a loss of precision of only a few percent is encountered. In many experiments of this type currently being practiced there are up to 4 times as many backcrosses as F_2 individuals scored and such designs will reduce efficiency by from 70 to 120 per cent.

(ii) *The North Carolina Expt. III*

As was shown in section 2(ii) (see table 5) irrespective of the genetical control, one needs to sample at least 20 F_2 parents. Little extra gain in efficiency is obtained from larger samples while efficiencies decline rapidly with less than 20 parents. If one chooses to use more than 20 F_2 parents then there is a proportional decrease in the number of replicates required so that the total experimental size remains approximately constant. Obviously it is not possible to reduce the number of replicates below 2, therefore above about $k = 50$ the minimum total experimental size will start to increase.

The only significant dependence of the structure on the parameters and genetical control thus lies in the degree of replication required, more replication being required for H than D , with low heritabilities and if the genes are linked in association.

4. RELATIVE EFFICIENCIES OF THE TWO DESIGNS

The minimum total experimental sizes necessary to achieve conventional 5 per cent significance for estimates of D and H are given in tables 4 and 5 for the N.C. Expt. III and in tables 2b and 3b for the F_2 /backcross experiment.

A comparison of these values for the two designs clearly indicates the far greater efficiency of the N.C. Expt. III. With a low heritability (0.2) it requires only one fifth the total experimental size for estimating D and only one seventieth for H . At high heritabilities, though less marked, the N.C. Expt. III is from two to 17 times as efficient for estimating D and H respectively.

These differences in total experimental size are not counter balanced by the increased effort required in the crosses needed to produce the appropriate families. Clearly with a natural inbreeder it is technically easier to produce F_2 's than crosses. However, as the figures above show, the F_2 /backcross design always requires as many crosses (and generally far more), than does the N.C. Expt. III, and so there can be no saving in effort at this stage.

5. EXPERIMENTAL SIZES REQUIRED

It is clear from tables 4 and 5, that in order to estimate D from the N.C. Expt. III no more than 600 completely randomised individuals are required even for heritabilities as low as 0.2. For H more than 1000 individuals are required for heritabilities below 0.4. Such a number is not large for a single experiment, but soon becomes unmanageable if many F_2 's are to be compared simultaneously.

However, it is often technically convenient to raise many organisms in plots, particularly crop plants which are to be assessed under normal agricultural practice. By working with plot means the heritabilities are effectively increased; thus, although the total number of individuals will increase (individual randomisation being the most efficient design), the number of replicated experimental units decreases as a function of plot size. Such a procedure is not however practical with the F_2 /backcross experiment.

With the F_2 /backcross design the size varies greatly with the genetic models and parameters to be estimated. It is generally not a practicable method of estimating dominance although one can successfully estimate additive variation with moderately sized experiments (1000) providing that the heritability is not less than 30 per cent.

6. DISCUSSION

In order to predict the properties of recombinant inbred lines derived from an F_2 , one requires information either on the additive genetic variation D or on D together with H (the dominance variation). Providing there is no linkage, epistasis or genotype-environment interaction the expected variance of the means of a set of inbred lines derived from an F_2 is D . Given that the distribution of line means is approximately normally distributed the proportion of derived lines exceeding any specified score can be predicted knowing D and the F_2 mean. Similarly, if dominance is present, use of the inverse dominance ratio will enable one to predict the likely score of the best inbred line to be obtained from inbreeding the F_2 .

If linkage and epistasis exist among the genes controlling the character in question the estimate of D obtained from early generations will be biased, but generally the bias will be precisely the same among the derived inbreds and hence will still be the appropriate predictor of inbred performance. Apart from the very real possibility that the ranking of genotypes varies over seasons or locations, genotype environmental interaction does not present any problems in prediction.

Thus in order to assess the future potential of generations derived from a number of inbred lines and hence to choose particular crosses to inbreed, one simply requires efficient estimates of D (and possibly H). The previous evidence clearly indicates that the N.C. Expt. III is always the most efficient approach to obtain these estimates. For every pair of inbred lines each experiment need involve no more than 20 F_2 individuals. The data may be recorded on individuals although in many practical situations replicated plots are more efficient. Promising crosses identified in this way should then be selfed for some 6 to 8 generations without selection, using the method of individual seed descent from some 100 or more F_2 individuals. At the end of the inbreeding programme the derived inbreds can then be assessed and selected in large replicated trials.

For many commercial traits there are optimal harvesting times which may vary with the genotype. Such situations can be considered as a special type of genotype environmental interaction. Again this poses no real problem since a given N.C. Expt. III can be raised with several replicate plots per family, two or more plots being randomly assigned to each one of the harvesting dates. Clearly if " D " shows a major interaction with harvesting date it is possible that different line crosses would be needed for each, but the predictions described above would resolve this problem before the inbreeding programme is initiated.

Clearly the number of pairs of crosses that can be examined in this way is limited by time and space and it is likely that ten or less would be the norm at any given site and year. The identification of this small subset of potential lines from a much larger collection of material requires a quite different approach and this is presently under investigation.

7. REFERENCES

- COMSTOCK, R. E., AND ROBINSON, H. F. 1952. Estimation of average dominance of genes. *Heterosis*, Chap. 30. Iowa State College Press.
- JINKS, J. L., AND PERKINS, J. M. 1969. The detection of linked epistatic genes for a metrical trait. *Heredity*, 24, 465-575.
- JINKS, J. L., AND PERKINS, J. M. 1972. Predicting the range of inbred lines. *Heredity*, 28, 399-403.
- KEARSEY, M. J. 1970. Experimental sizes for detecting dominance variation. *Heredity*, 25, 529-542.
- KEARSEY, M. J., AND JINKS, J. L. 1968. A general method of detecting additive dominance and epistatic variation for metrical traits. 1. Theory. *Heredity*, 23, 403-409.
- MATHER, K. M., AND JINKS, J. K. 1971. *Biometrical Genetics*. Chapman and Hall.
- POONI, H. S., AND JINKS, J. L. 1976. The efficiency and optimal size of triple test cross designs for detecting epistatic variation. *Heredity*, 36, 215-227.