Estimation of the true additive genetic variance $\sum_{i=1}^{k} d_i^2$ in the presence of linkage disequilibrium

H. S. Pooni
and J. L. Jinks

Department of Genetics, University of Birmingham, Birmingham B15 2TT, U.K.

Accurate estimates of the true additive genetic variance ($\sum_{i=1}^{k} d_i^2$) of a cross between two pure breeding varieties can be obtained from the additive genetic components of the first three ranks (D_1, D_2 and D_3) when the latter are biased by the presence of linkage. Additive genetic variances of the lower ranks are directly equatable with $\sum_{i=1}^{k} d_i^2$ because they incur minimal bias even when the predominating linkages are strong. More precise estimates of $\sum_{i=1}^{k} d_i^2$ are however obtainable from the asymptotic regression analysis or a weighted least squares analysis.

Estimates of $\sum_{i=1}^{k} d_i^2$ when obtained from 784 hierarchically derived F_7 families of the $V2 \times V12$ cross of *Nicotiana rustica* were observed to be considerably larger than the additive genetic variance displayed by the F_{13} inbreds of the same cross for all the characters that showed significant excess of repulsion linkages. These results lend support to our commonly held view that the prediction procedures generally underestimate the probability of successful recovery of superior recombinant inbreds.

INTRODUCTION

The effects of linkage on the expectations and estimates of additive genetic variance (D) are well documented (Jinks and Pooni, 1976, 1981, 1982, 1984; Kearsey, 1985; Mather and Jinks, 1982). In general, a linkage disequilibrium after summing over all pairs of loci, though not linkage *per se*, leads to an underestimation or overestimation of the expected additive genetic variance $\sum_{i=1}^{k} d_i^2$ for an excess of repulsion phase or coupling phase linkages, respectively (Jinks and Pooni, 1982). Thus, while the biased estimates of the additive genetic variance from the early generations of a cross may approximate closely with the hitherto biased heritable variance of the F_∞ inbreds, neither of these statistics provide the true magnitude of the additive genetic variance when linkage disequilibrium is significant. In this paper we explore the possibilities of obtaining linkage free estimates of the additive genetic variance $\sum_{i=1}^{k} d_i^2$ and indeed show that this is possible both in theory as well as in practice.

THEORY AND METHODS

According to Mather and Jinks (1982) the total additive genetic variance ($_D F_n$) in the nth genera-tion of selfing from a cross between two pure breeding varieties can be obtained from the summation $\sum_{r=1}^{n-1} (\frac{1}{2})^r D_r$, where r is the rank of the source of variability. In the absence of linkage $D_r = \sum_{i=1}^{k} d_i^2$ for each of the $r = 1$ to $r = n - 1$ ranks. However, when linkages are significant the expectations of the additive genetic variances vary between the ranks and

$$D_r = \sum_{i=1}^{k} d_i^2 \pm \frac{C}{R} \sum_{i=1, j=i+1}^{k} 2(1 - 2p_{ij})^r d_i d_j$$

where the linkage component $2(1 - 2p_{ij})^r d_i d_j$ is added for each pair of genes linked in coupling (C) and subtracted for each pair linked in repulsion (R) (for details see Jinks and Pooni, 1982). Therefore additive genetic variances (D) when estimated from different ranks of a hierarchical analysis have the following expectations.

$$D_1 = \sum_{i=1}^{k} d_i^2 \pm \frac{C}{R} \sum_{i=1, j=i+1}^{k} 2(1 - 2p_{ij}) d_i d_j$$

$$D_2 = \sum_{i=1}^{k} d_i^2 \pm \frac{C}{R} \sum_{i=1, j=i+1}^{k} 2(1 - 2p_{ij})^2 d_i d_j$$

$$D_3 = \sum_{i=1}^{k} d_i^2 \pm \frac{C}{R} \sum_{i=1, j=i+1}^{k} 2(1 - 2p_{ij})^3 d_i d_j \dots$$

It is clear from these expectations that the contributions of $\sum_{i=1}^{k} d_i^2$ to the additive genetic variance of each rank remains the same throughout. The coefficients of linkage, on the other hand, follow the geometric series and their contributions decrease rapidly as r increases. Thus, the estimates of D_r when $r \geqslant 5$ must approximate closely with $\sum_{i=1}^{k} d_i^2$, the true additive genetic variance of the cross. We shall therefore exploit this unique property of the rank variances and assume that the D_r of the lowest rank is equal to $\sum_{i=1}^{k} d_i^2$ (method 1).

The estimates of D_r from different ranks, however, may not differ significantly especially when r is large. Also, the larger sampling errors of the variance components may make the magnitude of $D_{(n-1)}$ smaller than $D_{(n-2)}$ or even $D_{(n-3)}$ when it is expected to take a larger value and vice versa. A more consistent estimate of D_r should therefore be obtained by the weighted least squares analysis and we shall use this technique to obtain a more accurate estimate of $\sum_{i=1}^{k} d_i^2$ (method 2).

Further, we can calculate the expected values of $(1-2p_{ij})$, $(1-2p_{ij})^2$, $(1-2p_{ij})^3$ etc. by substituting $(D_3 - D_2)/(D_2 - D_1)$ and in turn substitute these values as coefficients of the linkage disequilibrium component Δ in a model which allows E the environmental component, $\sum_{i=1}^{k} d_i^2$ and Δ to be estimated by weighted least squares (method 3). This procedure also allows us to obtain parallel estimates of $\sum_{i=1}^{k} d_i^2$ when the average value of $(1-2p_{ij})$ can be calculated by more than one method. However, the reliability of these estimates is completely dependent on the accuracy with which the average value of $(1-2p_{ij})$ is estimated.

Alternatively, we can vary the level of linkage disequilibrium (method 4). Many of these models, however, may fit the data and we shall select the most appropriate one by applying the following criteria.
(a) All parameters must be significant.
(b) The χ^2 of goodness of fit must take a non-significant value and individual $(O_i - E_i)^2/w_i$ values should be smaller than 2·0.
(c) The level of linkage disequilibrium must be only one or two steps higher than the one which makes the χ^2 significant.

Finally, when there is a preponderance of repulsion linkage the consecutive values of the rank variances follow *Mitscherlich's law* (1930). Thus the value of D_r steadily approaches its maxima of $\sum_{i=1}^{k} d_i^2$ as r becomes large. Therefore, we can estimate $\sum_{i=1}^{k} d_i^2$ as an asymptote by carrying out an asymptotic regression analysis (method 5). The procedural details of this analysis are available from Patterson (1956, 1960) and Snedecor and Cochran (1967) and will not be repeated here.

EXPERIMENTAL DESIGNS

We require at least three additive genetic variances of different ranks (say D_1, D_2 and D_3) to obtain a reasonably accurate estimate of $\sum_{i=1}^{k} d_i^2$. Traditionally they can be obtained from the pedigree families derived by self-pollination or sib mating (Jinks and Pooni, 1984; Virk et al., 1981; Cooke and Mather, 1962). However, in the absence of non-additive effects the hierarchical structure of the selfing or sib mating series must be taken to at least the fifth generation (F_5 or S_5) to obtain the estimates of D_1, D_2 and D_3 (Jinks and Pooni, 1982).

Alternatively, independent and unbiased estimates of D_1, D_2 and D_3 can also be obtained by extracting doubled haploids from the F_1, F_2 and F_3 generations of a selfing series. Another design that can provide the estimates of these components is the triple test cross. However, one needs to produce three triple test crosses (one each from the F_2, F_3 and F_4 generations) and that would be expensive in time and space, although they will also provide an excellent opportunity for obtaining the three corresponding estimates of $\sum_{i=1}^{k} h_i^2$ that is, H_1, H_2 and H_3.

Finally, if it is possible to cross the material easily then the F_3 and F_4 generations can be replaced by the progenies of the first and second cycles of randomly mated F_2's in which linkage disequilibrium declines by a factor of $(1 - p_{ij})$ with each cycle of random mating.

EXAMPLE

We demonstrate the practical applicability of our procedures by estimating $\sum_{i=1}^{k} d_i^2$ from hierarchically derived inbred lines of the cross between varieties 2 and 12 of *Nicotiana rustica*. A large sample (784) of these inbreds and 60 F_{13} lines obtained from the same cross by single seed descent were raised during the summer of 1982. Individual plants from the experiment were scored for heights at 4 and 6 weeks after planting in the field (H_4 and H_6), flowering time (FT), height at flowering time (HFT), leaf length (LL), leaf width (LW) and final height at the end of the season (FH). These data were initially analysed by Jinks and Pooni (1984) who also provide fuller details of the experiment and analytical procedures.

Hierarchical analysis of the 784 pedigree inbreds yields five mean squares. These and the relevant items of the analysis of variance of 60 F_{13}

inbred lines have been tabulated in table 1. These analyses were originally given by Jinks and Pooni (1984) and are reproduced here for reference only. The σ^2's associated with each of these mean squares have the following expectations when epistasis and genotype × environment interactions are non-significant.

Hierarchical lines $\qquad$ F_{13} lines

$\sigma_1^2 = \frac{1}{2}D_1$ $\qquad$ $\sigma^2 bF_{13} = \frac{1}{2}D_1 + \frac{1}{4}D_2 +$

$\sigma_2^2 = \frac{1}{4}D_2$ $\qquad$ $\dots (\frac{1}{2})^{12}D_{12} \simeq {}_DF_\infty$

$\sigma_3^2 = \frac{1}{8}D_3$ $\qquad$ $\sigma^2 bf =$ Interaction between

$\sigma_4^2 = \frac{1}{16}D_4 + \frac{1}{32}D_5$ $\qquad$ blocks and families

$\sigma_5^2 = \frac{1}{64}D_6 + E$ $\qquad$ $\sigma^2 wF_{13} = (\frac{1}{2})^{13}D_{13} + E$

For each character we estimate $\sum_{i=1}^{k} d_i^2$ from these statistics in the following ways. (a) We estimate, D_1, D_2, D_3 and D_4 ($=D_5=D_6=D_7$) from the mean squares of the hierarchical analysis and equate D_4 with the true additive genetic variance $\sum_{i=1}^{k} d_i^2$. (b) We fit different D_r's to the same mean squares following Jinks and Pooni (1984) and equate the weighted least squares estimate of D_4 with $\sum_{i=1}^{k} d_i^2$. (c) We estimate the average value $(1 - 2p_{ij})$ from the perfect fit solutions of D_1, D_2 and D_3 wherever possible and obtain the estimates of $\sum_{i=1}^{k} d_i^2$ by weighted least squares analysis. (d) We fit the E, Δ and $\sum_{i=1}^{k} d_i^2$ model sequentially to the five mean squares of the hierarchical families by varying the average value of $(1 - 2p_{ij})$ from $0 \cdot 05$ to $0 \cdot 50$ at a constant interval of $0 \cdot 05$ and the best model is chosen according to the criteria specified in section 2. (e) All the eight mean squares of table 1 are also subjected to sequential model fitting as

described above but two additional parameters $(\sigma^2 wF_{13}$ and $\sigma^2 bf)$ are allocated to account for the effects of the environmental and interaction components of F_{13} lines. (f) We estimate the asymptote $\sum_{i=1}^{k} d_i^2$ from an asymptotic regression analysis of D_1, D_2, D_3 and D_4 ($= D_5$).

These estimates and their square roots $(\sqrt{\sum_{i=1}^{k} d_i^2})$ are tabulated for each of the seven characters in table 2. The corresponding additive genetic variances of the F_{13} inbreds and their square roots are also given for comparison. These have been taken from table 4 of Jinks and Pooni (1984).

CONCLUSIONS

In this paper we have set out to demonstrate that it is possible to obtain unbiased estimates of the additive genetic variance $\sum_{i=1}^{k} d_i^2$ in the presence of linkage disequilibrium. These estimates can be reliably obtained from the first three ranks (D_1, D_2 and D_3) of the additive genetic variance that are extractable from the pedigree families of a selfing or sib mating series (see Jinks and Pooni, 1984; Virk et al., 1981; Cooke and Mather, 1962, for details). Because pedigree inbreeding is the lynch pin of a large number of breeding programmes this kind of information is often available and therefore can be used for estimating $\sum_{i=1}^{k} d_i^2$. This in turn can provide a second opinion about the potential of the breeding material at a half way stage of a breeding programme and thus can help the breeder in formulating a long term breeding strategy.

In theory, methods 1 and 2 provide estimates that are only close approximations to the true

Table 1 Relevant mean squares for the hierarchical and F_{13} inbred lines

Source	Item	d.f.	H_4	H_6	FT	HFT	LL	LW	FH	EMS
							Character			
Hierarchical inbreds	Between F_2 groups	96	1514·82‡	7141·36	3282·33	13640·71	386·91	471·09	24254·97	$\sigma_5^2\dagger + 8\sigma_4^2 + 16\sigma_3^2 + 32\sigma_2^2 + 64\sigma_1^2$
	Between F_3 groups/F_2	98	386·00	2203·19	1195·22	5741·09	119·21	126·90	7825·78	$\sigma_5^2 + 8\sigma_4^2 + 16\sigma_3^2 + 32\sigma_2^2$
	Between F_4 groups/F_3/F_2	196	183·85	841·78	319·87	1786·30	38·90	44·73	2186·85	$\sigma_5^2 + 8\sigma_4^2 + 16\sigma_3^2$
	Between F_7 families/F_4/F_3/F_2	392	73·73	346·80	153·58	756·85	17·49	19·63	1006·38	$\sigma_5^2 + 8\sigma_4^2$
	Within F_7 families	5347	19·10	91·36	33·83	276·09	8·04	8·23	287·46	σ_5^2
F_{13} lines	Between SSD families	59	534·85	3027·61	1246·12	7587·19	194·75	218·64	9942·26	$\sigma_{wF_{13}}^2 + 8\sigma_{bf}^2 + 16\sigma_{bF_{13}}^2$
	Blocks × families	59	30·61	122·44	30·76	314·23	7·39	10·10	274·18	$\sigma_{wF_{13}}^2 + 8\sigma_{bf}^2$
	Within families	816	17·14	69·70	28·85	237·50	8·47	9·13	255·44	$\sigma_{wF_{13}}^2$

† for definitions of σ^2's see text.
‡ Every "between families" mean square is significant when tested against the appropriate error.

Table 2 Estimates of the true additive genetic variance $\sum_{i=1}^{k} d_i^2$, the additive genetic variance (DF_{13}) of the F_{13} inbred lines and their square roots for the seven characters.

Source/method	H_4 1†	H_4 2	H_6 1	H_6 2	FT 1	FT 2	HFT 1	HFT 2	LL 1	LL 2	LW 1	LW 2	FH 1	FH 2
1. Lowest rank	72·85	8·54	340·59	18·46	159·68	12·64	641·07	25·32	12·59	3·55	15·25	3·91	958·61	30·96
2. W. L. S. estimates for the lowest rank	66·40	8·15	310·75	17·63	157·34	12·54	574·55	23·97	10·81	3·29	12·77	3·57	841·00	29·00
3. W.L.S. estimates using the calculated value of $(1-2p_{ij})$	—	—	—	—	—	—	589·85	24·29	—	—	—	—	—	—
4a. Serialised model fitting to five mean squares	106·88	10·34	313·67	17·71	219·58	14·82	582·58	24·14	10·81	3·29	12·68	3·56	824·71	28·72
4b. Serialised model fitting to eight mean squares	85·87	9·27	330·08	18·17	159·96	12·65	587·06	24·23	10·92	3·30	12·75	3·57	863·30	29·38
5. Asymptotic regression	—	—	714·91	26·74	188·44	13·72	749·60	27·38	11·80	3·44	13·70	3·70	869·70	29·49
Average		9·08		19·74		13·27		24·84		3·37		3·70		29·51
F_{13} inbreds	31·32	5·60	181·57	13·47	76·07	8·72	454·56	21·32	11·65	3·41	13·09	3·66	605·35	24·60

† 1 = additive genetic variance; 2 = square root of the additive genetic variance.

magnitude of $\sum_{i=1}^{k} d_i^2$. These estimates can, therefore, contain an appreciable amount of bias especially when the linkages are tight. The efficiencies of methods 3 and 5, on the other hand, are directly related to the precision of the estimates of D_1, D_2 and D_3. Thus they can also fail to provide realistic estimates of $\sum_{i=1}^{k} d_i^2$ when the estimates of the rank variances are unreliable. However, it is highly improbable that all of these methods will fail simultaneously for the same reason. Nevertheless until there has been further theoretical and experimental investigations of the relative reliabilities of different methods one is well advised to employ as many of them as possible to arrive at a consensus estimate of $\sum_{i=1}^{k} d_i^2$.

In practice it is reassuring that four of the five methods provide acceptable estimates of $\sum_{i=1}^{k} d_i^2$ for each character irrespective of the presence (for H_4, H_6, FT, HFT and FH) or absence (for LL and LW) of linkage disequilibrium. This shows that with the quality of data available from our experiments these four methods are capable of providing consistent estimates of the true additive genetic variance under all situations. However, method 3 has consistently failed to provide a realistic estimate of $\sum_{i=1}^{k} d_i^2$ for all characters except HFT and this can only be due to the low reliability of the estimates of D_1, D_2 and D_3. Method 5 on the other hand provides an unrealistic estimate for H_4 and its estimate for H_6 is twice as large as any other. Again, the source of this ambiguity is the negative correlation that exists between the estimates of rank variances for these characters.

These estimates further confirm that the true additive genetic variance of the $V2 \times V12$ cross is much larger than the genetic variance of the first cycle inbreds (F_{13} lines) for all characters except leaf length (LL) and leaf width (LW). Hence we have always underestimated the true inbreeding potential of this cross by a wide margin. We can, therefore, expect to extract inbreds which are even more superior and have a wider range of performances in the second and third cycles of inbreeding.

REFERENCES

COOKE, P. AND MATHER, K. 1962. Estimating the components of continuous variation. II. Genetical. *Heredity*, 17, 211–236.

JINKS, J. L. AND POONI, H. S. 1976. Predicting the properties of recombinant inbred lines derived by single seed descent. *Heredity*, 36, 253–266.

JINKS, J. L. AND POONI, H. S. 1981. Properties of pure breeding lines produced by dihaploidy, single seed descent and pedigree inbreeding. *Heredity*, 46, 391–395.

JINKS, J. L. AND POONI, H. S. 1982. Predicting the properties of pure breeding lines extractable from a cross in the presence of linkage. *Heredity*, 49, 265–270.

JINKS, J. L. AND POONI, H. S. 1984. Comparison of inbred lines produced by single seed descent and pedigree inbreeding. *Heredity*, 53, 299–308.

KEARSEY, M. J. 1985. The effect of linkage on additive genetic variance with inbreeding an F_2. *Heredity*, 55, 139–143.

MATHER, K. AND JINKS, J. L. 1982. *Biometrical Genetics*. 3rd Edn. Chapman and Hall, London.

MITSCHERCHLICH, E. A. 1930. *Die Bestimmung des Düngerbedürfnisses des Bodens*. Paul Parey, Berlin.

PATTERSON, H. D. 1956. A simple method for fitting an asymptotic regression curve. *Biometrics*, 12, 323–329.

PATTERSON, H. D. 1960. A further note on a simple method for fitting an exponential curve. *Biometrika*, 47, 177–180.

SNEDECOR, G. W. AND COCHRAN, W. G. 1967. *Statistical Methods*. Sixth Edn. Iowa State University Press, Iowa, U.S.A.

VIRK, D. S., POONI, G. S., INGRAM, N. R. AND JINKS, J. L. 1981. Analysis of induced quantitative variation. I. Hierarchical and triple test cross analysis. *Heredity*, 47, 317–326.