© 1981. The Genetical Society of Great Britain

PROPERTIES OF PURE-BREEDING LINES PRODUCED BY DIHAPLOIDY, SINGLE SEED DESCENT AND PEDIGREE BREEDING

J. L. JINKS and H. S. POONI

Department of Genetics, University of Birmingham, Birmingham B15 2TT Received 3.xii.80

SUMMARY

In the absence of differential selection the phenotypic distributions of the pure-breeding lines extractable from an F_1 hybrid are identical irrespective of whether they are derived by dihaploidy or single seed descent. This identity is not affected by the kinds of gene action, genic interaction and genotype × environment interaction contributing to the phenotypic differences. Linkage discussibly delaying the extraction of the dihaploids until the F_2 or F_3 generations. The genetic components of the means, variances, skewnesses and kurtoses of these phenotypic distributions can be estimated from the early generations of a cross.

1. Introduction

It is becoming commonplace to produce random samples of pure-breeding lines (F_{∞}) from the F_1 of a cross by dihaploidy or by single seed descent either alongside or as an alternative to their production by more conventional methods such as pedigree or mass inbreeding. This has led to speculation about the expected properties of the populations of pure-breeding lines produced by these new procedures. In this paper we have, therefore, summarised their expected phenotypic distributions for a variety of circumstances using the biometrical genetical notation of Mather and Jinks (1971).

2. EXPECTATIONS

(i) No epistasis and linkage equilibrium

In the simplest case of no epistasis and linkage equilibrium the phenotypic distributions of the means of the pure-breeding lines produced as dihaploids and by single seed descent are identical. The statistics describing these distributions have the following expectations:

mean m

variance
$$D = \sum_{i=1}^{k} d_i^2$$

standard error \sqrt{D}

skewness 0

kurtosis
$$\left[\left(\sum_{i=1}^{k} d_i^4 + 6 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} d_i^2 d_j^2 \right) / \left(\sum_{i=1}^{k} d_i^2 \right)^2 \right] - 3$$

The mean m can be estimated as the mean of the parents $(P_1 \text{ and } P_2)$ of the initial cross i.e., $m = \frac{1}{2}(\bar{P}_1 + \bar{P}_2)$. $D = \sum_{i=1}^k d_i^2$, the additive genetic component of the variances of the phenotypic distributions, can be estimated either from the variances (V) of the F_2 and first backcrosses $(B_1$ and B₂) of the initial cross as

$$D = 4V_{F_2} - 2(V_{B_1} + V_{B_2})$$

or from the variance of sums in an F₂ triple test cross (Kearsey and Jinks, 1968; Jinks and Perkins, 1970).

Kurtosis can similarly be estimated from the sums in an F₂ triple test cross. Kurtosis, however, will occur only if either a few gene loci contribute to the phenotypic distribution (k is small) or there are inequalities in the additive genetic effects at different loci $(d_i \neq d_i)$.

(ii) With epistasis but no linkage disequilibrium

In the presence of epistasis the phenotypic distribution of the means of the pure-breeding lines produced as dihaploids and by single seed descent are still identical. The expected statistics of these distributions are, however, modified as follows. (Jinks and Pooni, 1976 and Pooni, Jinks and Cornish, 1977)

mean

variance
$$D + I = \sum_{i=1}^{k} d_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} i_{ij}^2$$

standard error $\sqrt{D+I}$

skewness
$$6\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}d_id_ji_{ij}+6\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\sum_{s=i+1}^{k}i_{ij}i_{js}i_{is}$$

(for kurtosis see Pooni *et al.*, 1977). The mean, m now equals $\frac{1}{2}\bar{P}_1 + \frac{1}{2}\bar{P}_2 + 4\bar{F}_2 - 2\bar{B}_1 - 2\bar{B}_2$ while the mean of the two original parents P_1 and P_2 , becomes m+[i], where [i] is the component of the mean of the phenotypic distribution ascribable to epistasis between homozygous combinations. Similarly, I is the corresponding component of the variances. There is no way of estimating the variance, skewness and kurtosis from the early generations of a cross. However, in the presence of the i type of interaction the additive genetic variance and the skewness and kurtosis estimated from the sums of an F₂ triple test cross are a closer approximation to the corresponding statistics for the pure-breeding lines than those estimated from any other source (Pooni and Jinks, 1979; Pooni, Jinks and Cornish, 1977).

(iii) No epistasis and linkage disequilibrium

Linkage disequilibrium is the sole cause of differences between the phenotypic distributions of pure-breeding lines produced by dihaploidy and by single seed descent. Their means remain the same and equal m, which is also the midparental value, but the variances and standard errors contain different linkage components. For example, the variance of dihaploids derived from an F_1 is

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

where p_{ij} is the recombination frequency between the *i*th and *j*th pair of loci and the linkage component is added if the pair of genes are in coupling and subtracted if they are in repulsion. This variance is identical with the estimate of the additive genetic variance from an F_2 triple test cross (Jinks, 1981).

The variance of the corresponding lines produced by single seed descent is

$$\sum_{i=1}^{k} d_i^2 + C \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left(\frac{2(1-2p_{ij})}{1+2p_{ij}} \right) d_i d_j.$$

Because of the greater opportunities for recombination, the contribution of any pair of genes to the linkage component is smaller for single seed descent than for dihaploidy. But for both, the magnitude of the linkage component is determined not by linkage per se but by the balance of contributions from coupling and repulsion pairs, i.e., the linkage disequilibrium on summing over all pairs of linked loci.

The presence, magnitude and direction of a linkage disequilibrium is readily determined in the early stages of a cross. It requires only that the variances of the B_1 , B_2 and F_2 families are compared with the total variances of the L_1 , L_2 and L_3 generations of an F_2 triple test cross (Perkins and Jinks, 1970).

In the presence of a linkage disequilibrium the variance of the dihaploid lines will be the larger if there is an excess of coupling pairs while that of the lines produced by single seed descent will be the larger if there is an excess of repulsion pairs. They can, however, be made to approach one another closely by deriving the dihaploids not from the F_1 but from the F_2 or F_3 .

The samples of dihaploids derived in this way will have variances equal to

$$\sum_{i=1}^{k} d_{i}^{2} + C \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left[(1 - 2p_{ij}) d_{i} d_{j} + (1 - 2p_{ij})^{2} d_{i} d_{j} \right]$$

and

$$\sum_{i=1}^{k} d_{i}^{2} + C \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left[(1-2p_{ij})d_{i}d_{j}\delta + \frac{1}{2}(1-2p_{ij})^{2}d_{i}d_{j} + \frac{1}{2}(1-2p_{ij})^{3}d_{i}d_{j} \right]$$

respectively.

These variances are expected to differ very little from that within a sample of pure-breeding lines produced by single seed descent.

Neither the dihaploids nor the inbreds produced by single seed descent will give a phenotypic distribution which is skewed (unless epistasis is also present) irrespective of the magnitude and direction of any linkage disequilibrium. It can, however, produce kurtosis even in the absence of other causes such as a small number of gene loci or unequals additive genetic effects (section 2(i)). This is because a linkage disequilibrium in which there is an excess of either coupling or repulsion combinations will mimic a smaller

number of gene loci with unequal additive genetic effects. Since the magnitude of the disequilibrium will be greater in samples of pure-breeding lines produced by dihaploidy than in those produced by single seed descent, the magnitude of the kurtosis will similarly differ in the same direction. But like all effects of a linkage disequilibrium this difference can be minimised by extracting the dihaploids from F_2 or F_3 generations.

(iv) With epistasis and linkage disequilibrium

A linkage disequilibrium between pairs of loci displaying non-allelic interactions is the only situation in which the means (\bar{F}_{∞}) of the phenotypic distributions of the pure-breeding lines produced from an F_1 hybrid by dihaploidy and single seed descent are expected to be different and not to equal m. Their expected means are

$$m + C \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (1-2p_{ij})i_{ij}$$

for dihaploidy and

$$m + C \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left(\frac{1-2p_{ij}}{1+2p_{ii}} \right) i_{ij}$$

for single seed descent.

This situation can be diagnosed in the early generations of a cross simply by comparing the means of the B_1 , B_2 and F_2 families with those of the L_1 , L_2 and L_3 sets of families of an F_2 triple test cross (Jinks, 1978). Failure of the equalities $\bar{B}_1 = \bar{L}_1$, $\bar{B}_2 = \bar{L}_2$ and $\bar{F}_2 = \bar{L}_3$ will detect the presence of linked interacting genes. The direction of the inequality will indicate the likely direction of the deviation of the \bar{F}_∞ from m.

Because of the linkage disequilibrium the variances of the phenotypic distributions of the pure-breeding lines will differ according to whether they are derived by haploidy or by single seed descent (Section 2(iii)) and because the disequilibrium involves interacting genes these interactions will also contribute to the difference. The expected variances are

$$\sum_{i=1}^{k} d_i^2 + \frac{C}{-R} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left[2(1-2p_{ij})d_id_j \right] + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left[4p_{ij}(1-p_{ij})i_{ij}^2 \right]$$

for dihaploidy and

$$\sum_{i=1}^{k} d_{i}^{2} + C \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left[\frac{2(1-2p_{ij})}{(1+2p_{ij})} d_{i}d_{j} \right] + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left[\frac{8p_{ij}}{(1+2p_{ij})^{2}} i_{ij}^{2} \right]$$

for single seed descent.

The phenotypic distributions will also be skewed and kurtosed and the two distributions will differ for both.

All of these differences between the phenotypic distributions of purebreeding lines from the two sources can, of course, be minimised by extracting the dihaploids from the F₂ or F₃ generations (section 2(iii)).

3. Conclusions

It should be apparent from the special cases described in section 2 that the properties of the pure-breeding lines extractable from the same F_1 hybrid by dihaploidy or by single seed descent are identical except where there is a linkage disequilibrium. Other complications, for example epi-

stasis, which is illustrated (section 2(ii)), and genotype × environment interaction, which is not, alter the shape of the distributions but do so identically for pure-breeding lines from both sources. All the advantages for subsequent biometrical genetical analysis of having access to the extremes of the distributions of pure-breeding lines (Jayasekara and Jinks, 1976) apply equally to material produced by both procedures.

All of our expectations assume that there has been no unconscious or deliberate selection during the production of the pure-breeding lines. Because the procedures for producing pure-breeding lines by dihaploidy and by single seed descent differ in so many ways, any unconscious selection is likely to differ also and, therefore, to lead to differences between their phenotypic distributions. Deliberate selection, unless imposed differently by design on lines produced by the two procedures, should not, however, lead to differences between their phenotypic distributions.

Unless, therefore, there is good cause to suspect that unconscious selection has occurred during the derivation of pure-breeding lines by one or both procedures any difference between their phenotypic distributions must be due to a linkage disequilibrium and this can be minimised by deriving the dihaploids from the F₂ or F₃ generation. Since this guarantees almost no differences in the properties of the pure-breeding lines produced by dihaploidy, single seed descent and, for that matter, pedigree breeding, choice of the method of extraction can be based solely on technical considerations, time and cost. The reduced opportunities for imposing deliberate selection during the production of the pure-breeding lines which are offered by dihaploidy and single seed descent relative to the more conventional procedures can be offset by the greater efficiency of the delayed selection (Jinks, Jayasekara and Boughey, 1977; Jinks and Pooni, 1981).

4. References

JAYASEKARA, N. E. M., AND JINKS, J. L. 1976. Effect of gene dispersion on estimates of components of generation means and variances. Heredity, 36, 31-40.

JINKS, J. L. 1978. Unambiguous test for linkage of genes displaying non-allelic interactions for a metrical trait. Heredity, 40, 171-173.

JINKS, J. L. 1981. The genetic framework of plant breeding. Trans. Roy. Soc. (In press).

JINKS, J. L., AND PERKINS, J. M. 1970. A general method for the detection of additive, dominance and epistatic components of variation. III. F₂ and backcross populations. Heredity, 25, 419-429.

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. Comparative results of selection in the early and late stages of an inbreeding programme. Heredity, 46, 1-7.

JINKS, J. L., JAYASEKARA, J. E. M., AND BOUGHEY, H. 1977. Joint selection for both extremes of mean performance and environmental sensitivity to a macroenvironmental variable. II. Single seed descent. Heredity, 39, 345-355.

KEARSEY, M. J., AND JINKS, J. L. 1968. A general method of detecting additive, dominance and epistatic variation for metrical traits. I. Theory. Heredity, 23, 403-409.

MATHER, K., AND JINKS, J. L. 1971. Biometrical Genetics (2nd Edn.). Chapman and Hall, London.

PERKINS, J. M., AND JINKS, J. L. 1970. Detection and estimation of genotype-environmental, linkage and epistatic components of variation for a metrical trait. Heredity, 25, 157-177.

POONI, H. S., AND JINKS, J. L. 1979. Sources and biases of the predictors of the properties of recombinant inbreds produced by single seed descent. Heredity, 42, 41-48.

POONI, H. S., JINKS, J. L., AND CORNISH, M. A. 1977. The causes and consequences of non-normality in predicting the properties of recombinant inbred lines. Heredity, 38, 329-338.