Theoretical and Applied Genetics 47, 35-39 (1976)

On the Power of Experimental Designs for the Detection of Linkage between Marker Loci and Quantitative Loci in Crosses between Inbred Lines

M. Soller and T. Brody

Department of Genetics, The Hebrew University, Jerusalem (Israel)

A. Genizi

Department of Statistics, Agricultural Research Organization, Rehovot (Israel)

Summary. The power of experiments aimed at detecting linkage between a quantitative locus and a marker locus, both segregating in the backross or F_2 generation of a cross between two inbred lines, is examined. Given that the two lines are close to fixation for alternative alleles of both marker locus and quantitative locus, it is concluded that experiments involving a few thousand offspring should be able to detect close linkages involving quantitative loci (or groups of loci) having rather modest effects (i.e., that contribute, say, 1% of the total phenotypic variance in the F_2).

Inbred lines, even if from the same base population, can be expected to differ in some of the quantitative and marker loci segregating in the original base population; chances of such differences are greater if the inbred lines originate from different base populations. Pleiotropy aside, therefore, quantitative effects associated with marker loci segregating in crosses between inbred lines can be attributed to differences between the inbred lines in quantitative loci linked to the marker locus. Such experiments could provide information on the distribution along the genome of loci affecting quantitative traits, and on the physiological specificity of such loci. They may provide answers to such questions as: are the quantitative effects associated with a particular marker locus limited to a particular trait, or do they extend through the entire woof and warp of the organism? Do they affect specific aspects of particular traits, or are they diffuse and general in nature? The great hope presented by such experiments is that they may isolate quantitative effects specific enough to be followed genetically in a semi-qualitative manner. The purpose of this paper is to provide some notion of the size of the experiments required to detect such linkages.

Theory and Results

Let A_1 and A_2 denote the alleles at the locus affecting the quantitative trait. Within each of the three genotypic classes, A_1A_1, A_1A_2, A_2A_2, the quantitative character, X, is assumed to be normally distributed with variance σ^2 and means, d, h, and -d, respectively. At the marker locus the alleles are denoted by M_1 and M_2, and it is assumed that the three marker genotypes have distinguishable phenotypes. Initially we assume complete linkage (probability of recombination, r=0) and fixation of alternative alleles in the two lines: line 1, with genotype M_1A_1/M_1A_1; and line 2, with genotype M_2A_2/M_2A_2. Other assumptions are as in Jayakar (1970).

Table 1 shows the expected frequency of the various marker classes among the two sorts of backcross progenies, and among the F_2 of a cross between the two lines. As a result of the linked quantitative loci, the offspring marker classes are expected to differ from one another in their value for the quantitative trait (also shown in Table 1). We now consider the sample size that would be required to detect a difference, δ, between the marker classes with Type I error, $\alpha = 0.05$, and Type II error, $\beta = 0.10$. In many cases, we are interested in the difference between two marker classes. Table 2 shows some differences of interest and the appropriate

Table 1. Expected Value and Frequency of Marker Classes among Backcross and F_2 Progeny[1]

Type of cross	Marker class					
	M_1M_1		M_1M_2		M_2M_2	
	freq.	value	freq.	value	freq.	value
Backcross to line 1	1/2	d	1/2	h		
Backcross to line 2			1/2	h	1/2	-d
F_2	1/4	d	1/2	h	1/4	-d

[1] details in text

Table 2. Values, Associated Variance and Number of Offspring Required per Experiment for some Differences of Interest According to Dominance at the Quantitative Locus[1]

Cross	Difference	Value	Variance	Number of offspring required according to Dominance at the Quantitative Locus				
				h = -d	h = $\frac{-d}{2}$	h = 0	h = $\frac{d}{2}$	h = d
Backcross to line 1	$\overline{M_1M_1} - \overline{M_2M_2}$	d - h	$\frac{2\sigma_w^2}{n}$	525	934	2100	8400	∞
Backcross to line 2	$\overline{M_1M_2} - \overline{M_2M_2}$	d + h	$\frac{2\sigma_w^2}{n}$	∞	8400	2100	934	525
Combined backcross	$(\overline{M_1M_1} - \overline{M_1M_2}) + (\overline{M_1M_2} - \overline{M_2M_2})$	2d	$\frac{4\sigma_w^2}{n}$	2100	2100	2100	2100	2100
F_2	$\overline{M_1M_1} - \overline{M_2M_2}$	2d	$\frac{2\sigma_w^2}{n}$	1050	1050	1050	1050	1050

[1] Proportionate effect of quantitative locus = 0.282, Type I error = 0.05, Type II error = 0.10, other details in text

variances for testing these differences. With $n > 30$ offspring per marker class, this is clearly a normal-test situation and the appropriate test value (λ) satisfying the error requirements can be obtained directly from the table of ordinates of the normal curve by summing $x_\alpha + x_\beta = \lambda$. For the specified α and β errors, λ equals 3.24. Thus, if the expected difference between the two marker classes (or the expected value of the contrast) is equal to δ, then the α and β errors will be satisfied when

$$\frac{\delta}{k\sigma_w^2/n} = \lambda, \text{ i.e., } n = \frac{k(3.24)^2}{(\delta/\sigma_w)^2}$$

where σ_w is the within marker class standard deviation, and $\frac{k\sigma_w^2}{n}$ is the variance of the particular difference ($k = 2$ or 4 depending on the number of marker classes involved) (see Table 2).

The number of offspring per marker class (n) is thus a function of δ/σ_w; δ, in turn, is a function of d and h. Thus, n is related to the proportionate effect, $D = \frac{2d}{\sigma}$, of the quantitative locus involved. In the one-way backcross design, each experiment involves two offspring groups, so the total number of offspring required is 2n; in the two-way backcross and in the F_2 designs, the total number of offspring required will be 4n. (In the F_2 design, offspring heterozygous at the marker locus are not included in the final analysis.) Table 2 gives the total number of offspring required to detect a quantitative locus having a proportionate effect, $D_0 = 0.282$. Such a locus will contribute an amount of additive genetic variance equal to 1% of the total phenotypic variance in the F_2 generation. This seems a reasonable magnitude of effect to look for: it is higher than the values obtained by Falconer (1960) from an analysis of selection studies in the mouse and Drosophila, but less than many of the values obtained in more recent studies aimed at identifying the effect of individual quantitative loci (see, e.g., Spickett and Thoday, 1966; Law, 1966; Wehrhahn and Allard, 1965; Chai, 1968; Al-Murrani and Roberts, 1974). Thus, we are considering loci for which the difference between alternative homozygotes ($2d = 0.282\sigma_w$) is equal to a bit more than one-quarter of a phenotypic standard deviation, i.e., we are considering loci determining a difference of 200 kg. in yearly milk production, 5-6 kg. in 14 month liveweight of beef cattle, 6-7 eggs in annual production of layers, and 30-40 grams in 8-week body weight of broilers. Since n is inversely proportional to D^2, offspring numbers required to detect loci having proportionate effects, D, other than $D_0 = 0.282$, can be obtained simply by multiplying the values in Table 2 by $(\frac{0.282}{D})^2$.

Table 3 shows the differences of interest, and required sample numbers per experiment class when there is full dominance for the marker locus. When the dominant quantitative locus is linked to the dominant marker locus, the backcross design is more sensitive than the

Table 3. Values, Associated Variance and Number of Offspring Required per Experiment for some Differences of Interest, when there is Dominance at the Marker Locus[1]

Cross	Difference	Value	Variance	Number of offspring required according to Dominance at the Quantitative Locus				
				$h = -d$	$h = \frac{-d}{2}$	$h = 0$	$h = \frac{d}{2}$	$h = d$
Backcross to line 2	$\overline{M_1M_2} - \overline{M_2M_2}$	$d + h$	$\frac{2\sigma_w^2}{n}$	∞	8400	2100	934	525
F_2	$(\overline{M_1-}) - \overline{M_2M_2}$	$\frac{4}{3}d + \frac{2}{3}h$	$\frac{4\sigma_w^2}{3n}$	6300	2800	1584	1008	700

[1] Proportionate effect of quantitative locus = 0.282, Type I error = 0.05, Type II error = 0.10, other details in text

F_2 design. The situation is reversed in the absence of dominance, or when the recessive quantitative locus is linked to the dominant marker locus. For both designs, when $h = 0$ experiments involving a total of 2000 offspring or so would be of the required power. Consideration of Tables 2 and 3 shows that in most cases the F_2 design requires fewer offspring for equivalent power than the backcross design, so that in planning such experiments, the F_2 design is to be recommended.

In the F_2, in the absence of marker dominance, there are situations ($h \neq 0$, depending on h) when including the heterozygous marker class in the analysis would increase its power. In this case the normal-test is no longer appropriate, and a one-way analysis of variance based on three classifications would be used. Simple manipulations of standard textbook formulae (Scheffé, 1959) then yield

$$N = \frac{\Phi_k^2(k)}{(\xi/\sigma_w)^2}$$

where,

Φ_k is the non-central parameter for given α error and β error,

k is the number of marker genotype classes among the offspring,

$\xi = \Sigma r_i(\xi_i)^2 = 1/4\,(2d^2 + h^2)$,

r_i is the frequency of each marker genotype class,

ξ_i is the deviation of each marker genotype class from the mean, and

N is the total number of offspring over all marker genotype classes required for specified power.

For $D = 0.282$ and $h = 0$, $d/2$ and d, the total numbers of offspring required for the desired power are 1248, 1085 and 832, respectively. Thus, once $h > d/2$, including the heterozygous marker class in the analysis of the intercross design will increase the sensitivity of the experiment.

Effect of Recombination

For all the normal-test situations, it can readily be shown that the effect of recombination is to reduce the expected value of the contrast, previously equal to δ, to $(1-2r)\delta$. Hence, sample sizes for given power will have to be increased by $1/(1-2r)^2$. It is clear that, once r is greater than 0.15-0.20, the chances of detecting a single linked quantitative locus for experiments of reasonable magnitude drop precipitously. If a number of quantitative loci are located in the vicinity of the marker locus, the situation is more complex.

For the analysis of variance situation, with three marker classes, it can readily be shown that in the presence of recombination,

$$\xi = 1/4(1-2r)^2\,[2d^2 + h^2(1-2r)^2].$$

So in this case, when r and $h \neq 0$, the power of the experiment is reduced even further than in the normal-test situation.

Effect of Incomplete Fixation at the Marker and Quantitative Locus

If one or both of the two lines to be crossed are not at fixation for alternative alleles of the quantitative locus, it is evident that the differences between marker genotypes will be less than in the situation of complete fixation. In this the effect of incomplete fixation is analo-

gous to the effect of recombination. Letting a_{11} and a_{12} be allelic frequencies in line 1, and a_{21} and a_{22} the allelic frequencies in line 2, it can readily be shown that the values for the various contrasts will be reduced by $(a_{11}-a_{21})$. Thus, required sample sizes are increased by a factor of $1/(a_{11}-a_{21})^2$. Clearly it will be difficult to detect linkage unless the strains involved are close to fixation (say a_{11}, $a_{22} \geqslant 0.8$) for the quantitative loci. Similarly, if the two lines are not at fixation for the marker locus, three classes of marker genotypes are produced among the backcross offspring. Letting m_{11}, m_{12} and m_{21}, m_{22} be allelic frequencies in line 1 and line 2 respectively, numerical examples show that so long as m_{11}, m_{22}, a_{11}, $a_{22} \geqslant 0.8$, it will be more efficient to disregard the minor marker genotype among the backcross offspring and to rely on the major genotypes alone. In this case the values of the relevant contrasts are reduced by approximately $(m_{11}-m_{21})(a_{11}-a_{21})$, and sample sizes must be increased accordingly. Furthermore, additional offspring of the unused marker genotype will have to be produced. Thus, if $m_{11} = 0.8$, and assuming that only heterozygous F_1 offspring are used as parents of the backcross, there will be 0.10 of M_2M_2 offspring in the backcross that are not utilized in the analysis.

Once $M_{11} < 0.8$, it will generally pay to include the M_2M_2 offspring and carry out an analysis of variance. It is cumbersome to write a completely general expression, but for the case $a_{11} = a_{22}$, $m_{11} = m_{22}$, it can readily be shown that

$$\xi = \frac{1}{8}(m_{11}-m_{21})^2(a_{11}-a_{21})^2$$

$$[1+(m_{11}-m_{21})^2](1-2r)^2[2d^2+h(1-2r)^2].$$

Discussion

These results show that experiments of rather modest total size, technically quite feasible for most experimental and agricultural species, are capable of detecting relatively small quantitative effects under ideal conditions (no recombination between marker locus and quantitative locus and complete fixation of alternative alleles in the lines being crossed). It is clear, however, that as soon as the lines involved move any degree from complete fixation, the power of the test decreases very rapidly. For example, in the case of complete fixation and $h = 0$, we have (from Table 2) 2100 offspring required for the total experimental size, for the backcross design. If $a_{11} = 0.8$ and $a_{21} = 0.1$, this rises to 4286 offspring. If, in addition, $m_{11} = m_{22} = 0.9$, 6696 offspring are required, and if we have $r = 0.1$, the total number of offspring required for the desired power rises to 10,462.

Practically speaking, these results mean that experiments of the sort described are feasible only if the two lines differ sufficiently for their marker loci so that it is possible to choose parents out of each line for the initial cross and the backcross which are homozygous for the marker loci under consideration. Furthermore, it will be possible to detect only those linked quantitative loci for which alternative alleles are close to fixation in the two lines.

The genetic interpretation of quantitative effects detected in this manner is beyond the scope of this paper. McMillan and Robertson (1974) have discussed the difficulty of determining whether a quantitative effect associated with a particular marker locus is produced by a single quantitative locus, or by a group of loci. In some cases the physiological specificity of the quantitative effect may provide a clue to its genetic architecture. Alternatively, insight may be provided by the maximum likelihood procedures developed by Stewart (1969) and Elston and Stewart (1973a, 1973b). Generally, the ability to locate a quantitative locus with respect to its associated marker locus will be limited by the confounding of effects due to recombination with those due to incomplete fixation in the inbred lines.

The laboratory mouse is an obvious candidate for experiments of this sort, particularly with respect to quantitative behavioral traits. Total genetic map size is quite small (Henderson and Edwards, 1968), numerous inbred lines are available, and these can be screened for a host of morphological, biochemical and antigenic markers (see, e.g., Taylor, 1972). Our own particular interest lies in the possibility of applying these designs to domestic chicken populations. In addition to their economic and biological interest, chickens are suitable material for such experiments: inbred lines are available, numerous genetic markers have been identified, and a variety of quantitative traits are well-defined (including: growth rate, body conformation, mature size, egg production and its components, egg size and quality, semen quality and quantity, and agonistic and sexual behaviour). Although there are $2n = 78$ chromosomes and a total genetic map length of $\simeq 5000$ centimorgans, most of the genetic information appears to be concentrated in the six

largest chromosome pairs (Pollack, 1974). Briggs and Nordskog (1973), in a study of Fayoumi × Heavy breed crosses, showed that significant effects on 8-week body weight, age at first egg, and ten-day egg production could be assigned to the Z sex chromosome. This is the fifth largest chromosome and contains almost 10% of the total genome (Ohno, 1967). Also, the known morphological markers all appear to map in a few linkage groups (Soames, 1973). Pollack (1974) has found an average total chiasmata number of 28 for these six major chromosome pairs. It would probably not be far wrong to assume that 70% of the genetic information in the domestic chicken is included in a total map length of 1500 centimorgans. Each marker locus in poultry would thus cover about 2% of the genome, so that an experiment involving 10-15 markers would monitor 20-30% of the genome. This should be sufficient to give a fair notion of the usefulness and interest of the information that can be obtained by this kind of experiment.

The main technical problem in designing such experiments with poultry (and other species as well) is to locate or develop strains for crossing that differ sharply in their gene frequencies for quantitative loci and for marker loci. Clearly, in determining the final choice of strains for crossing, it will always be an advantage to have strains that differ in a maximum number of marker loci. For the quantitative loci the situation is more complex. As strains differ in more quantitative loci, the chances of finding a quantitative effect associated with a particular marker locus increase, but so does the likelihood that the effect itself will be due to a block of linked quantitative loci rather than a single locus. Nevertheless, in any trade-off between the probability of detecting a quantitative effect, and precision in interpreting an effect, once detected, it seems reasonable that in initial experiments of this sort the major emphasis should be on maximizing the expected number and magnitude of marker-associated quantitative effects. The task of achieving a more precise interpretation of the effects uncovered, if any, can be left to later experiments.

Literature

Al-Murrani, W.K.; Roberts, R.C.: Genetic variation in a line of mice selected to its limit for high body weight. Anim. Prod. 19, 273-289 (1974)

Briggs, D.M.; Nordskog, A.W.: Influence of sex chromosomes on laying house performance of chickens. Brit. Poultry Sci. 14, 403-412 (1973)

Chai, L.K.: Analysis of quantitative inheritance of body size in mice. V. Effects of small number of polygenes on similar genetic background. Genetical Res. 11, 239-246 (1968)

Elston, R.C.; Stewart, J.: Biometrical genetics with one or two loci: The inheritance of physiological characters in mice. Genetics 73, 675-693 (1973a)

Elston, R.C.; Stewart, J.: The analysis of quantitative traits for simple genetic models from parental, F_1 and backcross data. Genetics 73, 695-711 (1973b)

Falconer, D.L.: Introduction to Quantitative Genetics. Edinburgh: Oliver and Boyd, 1960

Henderson, S.A.; Edwards, R.G.: Chiasma frequency and maternal age in mammals. Nature 218, 22-28 (1968)

Jayakar, S.D.: On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. Biometrics 26, 451-464 (1970)

Law, C.N.: The location of genetic factors affecting a quantitative character in wheat. Genetics 53, 487-498 (1966)

McMillan, I.; Robertson, A.: The power of methods for the detection of major genes affecting quantitative traits. Heredity 32, 349-356 (1974)

Ohno, S.: Sex Chromosomes and Sex-linked Genes. Berlin: Springer, 1967

Pollack, D.: Personal communication (1974)

Scheffé, H.: The Analysis of Variance. New York: Wiley 1959

Soames, R.G.: Linkage relationships in Domestic Fowl. J. of Heredity 64, 217-221 (1973)

Spickett, S.G.; Thoday, J.M.: Regular responses to selection. 3. Interaction between located polygenes. Genetical Res. 7, 96-121 (1966)

Stewart, J.: Biometrical genetics with one or two loci. I. The choice of a specific genetic model. Heredity 24, 211-224; II. The estimation of linkage. Heredity 24, 225-238 (1969)

Taylor, B.A.: Genetic relationships between inbred strains of mice. J. of Heredity 63, 83-86 (1972)

Wehrhahn, C.; Allard, R.W.: The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. Genetics 51, 109-119 (1965)

Received March 4, 1975
Communicated by A. Robertson

Dr. M. Soller
Dr. T. Brody
Department of Genetics
The Hebrew University
Jerusalem (Israel)

Dr. A. Genizi
Department of Statistics
Agricultural Research Organization
Rehovot (Israel)