Theor Appl Genet (1988) 75:857–862

The use of replicated progenies in marker-based mapping of QTL's*

N. M. Cowen

United AgriSeeds, Inc., P.O. Box 4011, Champaign, IL 61820, USA

Received May 18, 1987; Accepted September 7, 1987
Communicated by A. L. Kahler

Summary. Seven types of progeny are described which can be used in detection of linkage between marker loci and quantitative trait loci (QTL) in a cross between two inbred lines. Three types of progeny: recombinant inbred lines (RI); doubled haploid lines (DH); and S_1 lines can be used to detect linked main effects, d. DH and RI lines can be used to detect smaller effects than S_1 lines. However, S_1 lines can also be used to detect within-population dominance effects, h. The smallest d detectible is in the range of $\frac{1}{2}$ to $\frac{1}{12}$ the size of the corresponding $LSD_{(0.05)}$ for the quantitative trait, using 100 lines and 6 replicates. The smallest h detectible is 3–4 times this size. Four types of progeny can be used to detect differences in the dominance behavior of alleles within the population relative to an allele in another inbred line (P_4): DH lines × P_4; RI lines × P_4; either $F_2 \times P_4$ or S_1 lines × P_4; and progeny generated by crossing $(F_1 \times P_3) \times P_4$. Dominance differences in the range of $1\frac{1}{4}$ to $\frac{1}{6}$ the size of the corresponding $LSD_{(0.05)}$ are routinely detectible using 100 lines and 6 replicates. Increasing the numbers of progeny evaluated or the number of replicates allows for the detection of relatively smaller linked effects.

Key words: RFLP – Genetic markers – Genetic mapping – Experimental precision

Introduction

Attempts to resolve the underlying basis for variability in quantitative or polygenic traits into the effects of individual loci using linkage to markers were put on a firm theoretical basis by Thoday (1961). Since that time, numerous contributions to the theory of marker-based detection of quantitative trait loci (QTL's) have been made (McMillan and Robertson 1974; Soller et al. 1976; Soller et al. 1979; Soller and Genizi 1978). With the increase in the numbers of markers available for mapping, made possible through the use of restriction fragment length polymorphisms (RFLP's) and other detectible DNA polymorphisms, the area has recently received considerable attention (Soller and Beckman 1983; Beckman and Soller 1983; Botstein et al. 1980; Burr et al. 1983; Evola et al. 1986; Ellis 1986).

In most experimental designs aimed at the detection of linkage between a set of markers and QTL's, individuals of the population of interest are scored for the set of markers and evaluated for quantitative traits of interest. The primary exception to this is the use of recombinant inbred lines (Ellis 1986), the use of which is limited by the number of populations for which lines are available. The use of either individuals or recombinant inbred lines in marker-based detection of QTL's are based on a theory developed for animal systems. In contrast, numerous types of progenies, which can be replicated for evaluation, can be developed in relatively few generations in most plant systems. The purpose of this article is to outline the use of seven types of progenies in the detection of linked main and dominance effects in a population obtained by crossing two inbred lines.

The reference population

The reference population considered is the F_2 of a cross between two inbred lines, each of which has a

* Contribution of United AgriSeeds, Inc.

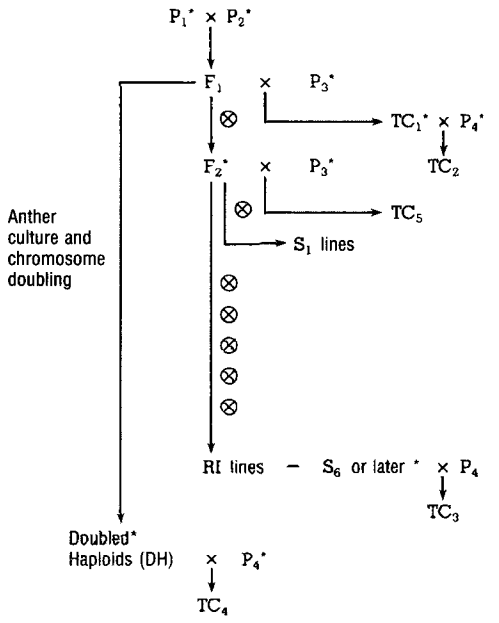
different allele at the marker locus and also at a linked QTL. The approach used to develop both the population and the progenies for evaluation is outlined in Fig. 1. The two parents of the population, P_1 and P_2, are genetically $M_1M_1A_1A_1$ and $M_2M_2A_2A_2$, respectively.

The recombination frequency between the marker and the QTL is r. The F_1 of the cross generates four possible gametes: M_1A_1, M_2A_2, M_1A_2 and M_2A_1 with frequencies $\frac{1}{2}(1-r)$, $\frac{1}{2}(1-r)$, $\frac{1}{2}r$ and $\frac{1}{2}r$, respectively. These gametes unite at random to form the F_2 resulting in nine possible genotypes, the frequencies of which are obtained as sums of products of the frequencies of the uniting gametes.

Types of progenies considered

Seven different progenies are considered for evaluation (Fig. 1). S_1 lines are generated by collection of seed resulting from self-pollination of individual F_2 plants. If S_1 lines are to be used in detection of linkage, then the individual F_2 plants, or a bulk of S_1 individuals within each family, will be scored for the marker. S_1 lines are useful in detection of linked main effects and within-population dominance effects as will be shown in a later section. Recombinant inbred lines, which for most purposes are S_6 or later lines (resulting from self pollination of F_7 individuals), can be scored directly for the markers with minimal sampling, and are useful in detecting linked main effects. Doubled haploids generated from the F_1 of the cross can be used exactly as the recombinant inbred lines.

Other types of progeny need to be generated to detect differences in the dominance behavior of alleles within the population relative to an allele in an unrelated inbred. The inbred, designated P_4 in Fig. 1, is considered to be unrelated to P_1 and P_2 and was chosen because it carried a different allele at the marker locus, M_4 (the genotype of P_4 is $M_4M_4A_4A_4$). In corn, P_1, P_2, and P_3 would probably be from the same heterotic group, while P_4 would represent a different heterotic group. The progenies used to evaluate differences in dominance behavior relative to A_4 are obtained by crossing either the doubled haploid lines (TC_4), the recombinant inbred lines (TC_3) or either F_2 individuals or S_1 lines (TC_5) to P_4. Additionally, a procedure which Stadler (1944) termed "gamete selection" can be carried out by crossing the F_1 to an inbred P_3. The resulting individuals (TC_1) are scored for the marker, crossed to inbred P_4 (TC_2) and are selfed. TC_2 families can also be used to detect linked dominance differences.



* Indicates that individual or line is scored for the markers

TC_2, TC_3, TC_4, and TC_5 all represent families rather than individuals

Fig. 1. Various progenies which may be useed in mapping QTLs in a population obtained by crossing two inbred lines

Evaluation of progenies

Evaluation of various progenies for the quantitative traits of interest can be accomplished using any appropriate design which allows for precise, unbiased estimates of progeny means. For the remainder of this article, progenies are considered to be evaluated in a randomized complete block design. The progenies are considered a random sample of that progeny type from the reference population.

The marker locus

The marker alleles are considered to be codominant and inherited in a Mendelian manner. The markers may be isozymes, RFLP's, other molecular markers, or morphological markers.

Expectations of various marker genotypes for the quantitative trait in the progenies considered

The genotypic values of individuals with genotype A_1A_1, A_1A_2, A_2A_2 are d, h, and $-d$, respectively. Let N_{M1M1}, N_{M1M2} and N_{M2M2} be numbers of lines either

genotypically M_1M_1, M_1M_2, and M_2M_2, or in the case of S_1 lines, the number of lines obtained by selfing F_2 individuals of the various genotypes. The expected ratio of $N_{M1M1}:N_{M1M2}:N_{M2M2}$ in the S_1 lines is 1:2:1.

The mean of all S_1 lines obtained from F_2 individuals which are M_1M_1 genetically is:

$$M_1M_1(S_1) = (1-r)^2 d + r^2(-d) + 2r(1-r)h$$
$$= (1-2r)d + 2r(1-r)h,$$

also

$$M_1M_2(S_1) = r(1-r)d + r(1-r)(-d) + [(1-r)^2 + r^2]h$$
$$= [(1-r)^2 + r^2]h$$

and

$$M_2M_2(S_1) = (2r-1)d + 2r(1-r)h.$$

Main effects are estimated as the difference between the means of the two homozygous marker classes:

$$M_1M_1(S_1) - M_2M_2(S_1) = 2(1-2r)d.$$

Dominance effects are estimated as the difference between the mean of the heterozygous marker class and the average of the homozygous marker classes:

$$M_1M_2(S_1) - \tfrac{1}{2}[M_1M_1(S_1) + M_2M_2(S_1)] = (1-2r)^2 h.$$

Both recombinant inbred lines (RI) and doubled haploid lines (DH) can be used to estimate main effects. However, they differ in the proportion of lines for any marker genotype which is recombinant. With DH lines obtained from the F_1 of the cross, the ratio of nonrecombinants to recombinants is $(1-r):r$ for any marker class, where r is the map distance between marker and QTL in Morgans. With RI lines, obtained through self pollination, the ratio of nonrecombinant to recombinant is $(1-R):R$ for any marker class, where $R = 2r/(1+2r)$ (Haldane and Waddington 1931). The expected ratio of M_1M_1 lines to M_2M_2 lines is 1:1 for both RI and DH lines.

The mean of all DH lines with marker genotype M_1M_1 is:

$$M_1M_1(DH) = (1-r)d + r(-d)$$
$$= (1-2r)d,$$

while the mean of all RI lines with marker genotypes M_1M_1 is:

$$M_1M_1(RI) = (1-R)d + R(-d)$$
$$= (1-2R)d.$$

The mean of all DH lines with genotype M_2M_2 is:

$$M_2M_2(DH) = (1-r)(-d) + rd$$
$$= (2r-1)d$$

and

$$M_2M_2(RI) = (1-R)(-d) + Rd$$
$$= (2R-1)d.$$

Main effects for both DH and RI lines are estimated as the difference between the means of the two marker classes. Thus, for DH lines, main effects are estimated as:

$$M_1M_1(DH) - M_2M_2(DH) = (1-2r)d - (2r-1)d$$
$$= 2(1-2r)d.$$

While for RI lines, main effects are estimated as:

$$M_1M_1(RI) - M_2M_2(RI) = (1-2R)d - (2R-1)d$$
$$= 2(1-2R)d$$
$$= 2[(1-2r)/(1+2r)]d.$$

As shown in Fig. 1, TC_3 and TC_4 families are generated by crossing RI and DH lines to P_4, respectively. The parent P_4 contributes an identical gamete (M_4A_4) to each individual in either TC_3 or TC_4 families, thus, differences in the means of these families are attributable to differences in the interaction of alleles A_1 and A_2 with A_4.

The genotypic value of the genotypes A_1A_4 and A_2A_4 are designated h_{14} and h_{24}, respectively. The mean value of all TC_3 families which are genetically M_1M_4 at the marker locus is:

$$M_1M_4(TC_3) = (1+2r)^{-1} h_{14} + 2r(1+2r)^{-1} h_{24},$$

while the mean value of all TC_4 families which are M_1M_4 at the marker locus is:

$$M_1M_4(TC_4) = (1-r)h_{14} + r h_{24}.$$

The mean value of all TC_3 and TC_4 families which are genetically M_2M_4 at the marker locus are:

$$M_2M_4(TC_3) = (1+2r)^{-1} h_{24} + 2r(1+2r)^{-1} h_{14}$$

and

$$M_2M_4(TC_4) = (1-r)h_{24} + r h_{14},$$

respectively. Differences in the dominance behavior of alleles A_1 and A_2 in combination with A_4 are detected as the differences between the means of lines genetically M_1M_4 and M_2M_4 for the marker. For TC_3 families, dominance differences are estimated as:

$$M_1M_4(TC_3) - M_2M_4(TC_3) = (1-2r)(1+2r)^{-1} h_{14}$$
$$+ (2r-1)(1+2r)^{-1} h_{24}$$
$$= (1-2r)(1+2r)^{-1}(h_{14} - h_{24}).$$

While for TC_4 families they are estimated as:

$$M_1M_4(TC_4) - M_2M_4(TC_4) = (1-2r)h_{14} + (2r-1)h_{24}$$
$$= (1-2r)(h_{14} - h_{24}).$$

The gametic output for the QTL of F_2 individuals (or S_1 line derived from it) which at the marker locus carry alleles M_1M_1 is:

$(1-r)\,A_1 + r\,A_2$.

The gametic output for the QTL of F_2 individuals (or S_1 line derived from it) which at the marker locus carry alleles M_1M_2 is:

$\frac{1}{2}(A_1 + A_2)$.

The gametic output for the QTL of F_2 individuals (or S_1 line derived from it) which at the marker locus carry alleles M_2M_2 is:

$r\,A_1 + (1-r)\,A_2$.

The genetic structure of testcrosses to P_4 of F_2 individuals (or S_1 line derived from it) which are M_1M_1, M_1M_2, and M_2M_2 for the marker are:

$(1-r)\,A_1A_4 + r\,A_2A_4$,

$\frac{1}{2}(A_1A_4 + A_2A_4)$,

and

$r\,A_1A_4 + (1-r)\,A_2A_4$,

respectively. The mean genotypic values of each of these 3 testcrosses are: $(1-r)\,h_{14} + r\,h_{24}$, $\frac{1}{2}(h_{14} + h_{24})$, and $r\,h_{14} + (1-r)\,h_{24}$, respectively.

Differences in the dominance behavior of alleles A_1 and A_2 in combination with A_4 are detected as differences between the means of TC_5 families of F_2 individuals (or S_1 line derived from it) with marker genotypes M_1M_1 and M_2M_2:

$$\begin{aligned} M_1M_1(TC_5) - M_2M_2(TC_5) &= [(1-r)\,h_{14} + r\,h_{24}] \\ &\quad - [r\,h_{14} + (1-r)\,h_{24}] \\ &= (1-2r)(h_{14} - h_{24}). \end{aligned}$$

The generation of TC_2 families is a two-step process. Gametes from the population are sampled by crossing the F_1 with the inbred P_3 which is genetically $M_3M_3A_3A_3$. These individuals, termed TC_1, are then crossed to P_4 to generate TC_2 families for evaluation. The genotypic values of TC_2 individuals with genotype A_1A_4, A_2A_4, and A_3A_4 are h_{14}, h_{24}, and h_{34}, respectively. The mean genotypic value of all TC_2 families resulting from the cross of TC_1 individuals with genotype M_1M_3 is:

$M_1M_3(TC_2) = \frac{1}{2}(1-r)\,h_{14} + \frac{1}{2}r\,h_{24} + \frac{1}{2}h_{34}$,

while the mean of all TC_2 families from M_2M_3, TC_1 individuals is:

$M_2M_3(TC_2) = \frac{1}{2}(1-r)\,h_{24} + \frac{1}{2}r\,h_{14} + \frac{1}{2}h_{34}$.

Differences in the dominance behavior of allele A_1 and A_2 in combination with A_4 are estimated as the difference between the means $M_1M_3(TC_2)$ and $M_2M_3(TC_2)$:

$$\begin{aligned} M_1M_3(TC_2) - M_2M_3(TC_2) &= \tfrac{1}{2}(1-2r)\,h_{14} + \tfrac{1}{2}(2r-1)\,h_{24} \\ &= \tfrac{1}{2}(1-2r)(h_{14} - h_{24}). \end{aligned}$$

The expectation of the difference for TC_2 families is exactly half the expectation of the difference for TC_4 families.

Examples of the size of detectable genetic effects

Some examples of the size of genetic effect which can be detected with a given linkage relationship, progeny type, number of replicates, and number of progeny can be very informative, particularly in relation to the size of the $LSD_{(0.05)}$ from the same experiment. The significance of any of the contrasts used to detect linked effects can be measured using either a *t*-test or an *F* test, which should give identical results. If multiple contrasts are going to be made with the same set of progeny, this is most easily accomplished by using a set of non-orthogonal contrasts in the analysis of variance, each of which is tested with the error mean square, or the entry × environment interaction mean square if the progeny are evaluated in multiple environments.

Using a set of 100 RI lines equally divided between the two marker classes (i.e., 50 M_1M_1 and 50 M_2M_2 lines) and replicated 6 times, the critical value for a *t*-test of linked main effects would be:

$t_{(0.05,\,df=495)}\sqrt{[(2 \times \text{MS entries})/300]}$.

The LSD from this experiment is:

$t_{(0.05,\,df=495)}\sqrt{[(2 \times \text{MS entries})/6]}$.

The mean square for source of variation entries is used both in the LSD and in the critical value for the *t*-test because entries are considered a random effect in the model (Cowen 1986). The critical value for the *t*-test is $0.160\sqrt{\text{MS entries}}$, while the LSD is $1.132\sqrt{\text{MS entries}}$ or roughly 7 times as large. The absolute size of the genetic effect detected as significant depends to a great extent on r. Dividing the critical value for the *t*-test by the coefficient on d from the contrast (see Table 1) gives a critical value for d for which we can examine the effects of changes in r. For example, with $r = 0.05$ the critical value is $0.098\sqrt{\text{MS entries}}$, while for $r = 0.40$ the critical value is $0.72\sqrt{\text{MS entries}}$, more than 7 times as large.

As shown, within population dominance effects (h) are detectable with S_1 lines. Using 100 lines and 6 replicates, with the frequencies of the various marker classes equal to their expectations (i.e., 1:2:1) the critical value for a *t*-test of the dominance contrast

Table 1. The coefficient in r of the genetic effects in the expectations of contrasts between marker classes for 7 progeny types

Progeny type	Genetic effect		
	Within population		
	Average d	Dominance h	$h_{14} - h_{24}$
DH	$2(1-2r)$		
RI	$2(1-2r)(1+2r)^{-1}$		
S_1	$2(1-2r)$	$(1-2r)^2$	
TC_2			$\frac{1}{2}(1-2r)$
TC_3			$(1-2r)(1+2r)^{-1}$
TC_4			$(1-2r)$
TC_5			$(1-2r)$

would be:

$$t_{(0.05,\,df=495)} \sqrt{[(2 \times \text{MS entries})/300]}\,.$$

This is identical to the value shown for d, however, the critical value divided by the coefficient on h for the expectation of the contrast (Table 1) takes values $0.198\sqrt{\text{MS entries}}$ and $4.00\sqrt{\text{MS entries}}$ for $r = 0.05$ and 0.40, respectively. The $LSD_{(0.05)}$ is $1.132\sqrt{\text{MS entries}}$ in this instance as well, thus, these effects are 0.17 and 3.5 times the corresponding LSD, respectively.

Using 100 TC_3 lines and 6 replicates, where the frequencies of the 2 marker classes are identical the critical value for the *t*-test of the contrast would be identical to that shown for the other contrast discussed. Adjusting the critical value for the coefficient in the expectation of the contrast (Table 1) gives critical values of $0.196\sqrt{\text{MS entries}}$ and $1.44\sqrt{\text{MS entries}}$ for $r = 0.05$ and 0.40, respectively. These values are approximately 0.17 and 1.25 times the size of the corresponding $LSD_{(0.05)}$.

Since the critical values in the *t*-test for all progeny depend on the numbers of progeny in either 2 or 3 possible marker classes, decreases in the numbers of progeny in a particular class are offset by increases in another, which minimizes the effects of sampling on the critical value. Increasing either the number of progeny and/or the number of replicates will allow detection of relatively smaller effects. The critical value for any combination of the number of progeny and replicates used, and progeny type, can be calculated with some knowledge of the possible linkage relationship and the effects of sampling.

Discussion

Three types of linked effects are detectible with the progeny described above. They are main effects, within population dominance effects, and dominance effects relative to an allele in an unrelated inbred line.

Linked main effects can be evaluated using S_1, DH and RI lines. Main effects are most easily detected using DH lines, which are only slightly more efficient than RI lines. Both DH and RI lines are better than S_1 lines. RI lines, unless already available, take considerably longer to develop than S_1 or DH lines. S_1 lines also have the advantage of allowing detection of linked, within population dominance effects.

Differences in the dominance behavior of alleles in the population in combination with an allele in an unrelated inbred ($h_{14} - h_{24}$) are detectible with either TC_2, TC_3 or TC_4 lines. TC_3 and TC_4 lines provide the easiest detection of these effects; TC_2 lines are useful in detecting effects approximately twice the size detectible with TC_3 or TC_4 lines.

The detection of dominance differences using TC_2 lines is independent of the choice of P_3. Thus either P_2 or P_1 can be substituted for P_3 (thus making TC_1 a backcross population) without affecting either the precision of, or the expectation of, the difference $M_1M_3(TC_2) - M_2M_3(TC_2)$.

Ellis (1986) examined the use of recombinant inbred lines in marker based mapping of QTL's and stated "RFLP markers can really only be used to follow the segregation of reasonably closely linked genes where the segregating alleles confer very different phenotypes (i.e., classical morphological characters). It is unlikely that they could be useful in the genetical analysis of quantitative characters." However, using a well populated genetic map, progeny which can be easily replicated, and a proper experimental technique for field trials, it is possible to detect segregating alleles with effects much smaller than the LSD. Hence, it is clear that numerous alleles of this nature affecting the same quantitative trait can be detected.

In all of the approaches outlined in this article, it appears that fewer individuals need to be scored for the marker, to detect effects of a given size, than in any approach previously described (Soller and Beckman 1983). Direct comparisons of these proposed approaches with previously described approaches will be dealt with in a manuscript in preparation. This may reduce the cost of marker based mapping of QTL's considerably (Beckman and Soller 1983).

References

Beckman JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor Appl Genet 67:35–43

Botstein D, White R, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Burr B, Evola SV, Burr FA, Beckman JS (1983) The application of restriction fragment length polymorphism to plant breeding. In: Setlow JK, Hollander A (eds) Genetic engineering principles and methods, vol 5. Plenum Press, New York, pp 45–59

Cowen NM (1986) Selection theory for selfed progenies. Theor Appl Genet 73:182–189

Ellis THN (1986) Restriction fragment length polymorphism markers in relation to quantitative characters. Theor Appl Genet 72:1–2

Evola SV, Burr FA, Burr B (1986) The suitability of restriction fragment length polymorphisms as genetic markers in maize. Theor Appl Genet 71:765–771

Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

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

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34:47–55

Soller M, Genizi A, Brody T (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47:35–39

Soller M, Brody T, Genizi A (1979) The expected distribution of marker-linked quantitative effects in crosses between inbred lines. Heredity 43:179–180

Stadler LS (1944) Gamete selection in corn breeding. J Am Soc Agron 36:988–989

Today JM (1961) Location of polygenes. Nature 191: 368–370