Theor Appl Genet (1990) 80:205–208

THEORETICAL AND APPLIED GENETICS

© Springer-Verlag 1990

Marker-based mapping of quantitative trait loci using replicated progenies

M. Soller [1] **and J. S. Beckmann** [2]

[1] Department of Genetics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
[2] Department of Plant Genetics and Breeding, Agricultural Research Organization, The Volcani Center, Bet Dagan 50250, Israel

Received January 19, 1990; Accepted February 9, 1990
Communicated by E. J. Eisen

Summary. When heritability of the trait under investigation is low, replicated progenies can bring about a major reduction in the number of individuals that need to be scored for marker genotype in determining linkage between marker loci and quantitative trait loci (QTL). Savings are greatest when heritability of the trait is low, but are much reduced when heritability of the quantitative trait is moderate to high. Required numbers for recombinant inbred lines will be greater than those required for a simple F_2 population when heritabilities are moderate to high and the proportion of recombination between marker locus and quantitative trait locus is substantial.

Key words: Quantitative trait locus (QTL) – Genetic markers – Mapping QTL – Recombinant inbred lines – Doubled haploids

Introduction

Theoretical studies of marker-based mapping of quantitative trait loci (Soller et al. 1976; Soller and Beckmann 1983; Beckmann and Soller 1986; Lebowitz et al. 1986) and experimental investigations based upon them (Edwards et al. 1987; Kahler and Wehrhahn 1986; Nienhuis et al. 1987; Paterson et al. 1988; Weller et al. 1988) have generally considered individual plants in simple F_2 or backcross populations. Such analyses are limited in that they cannot estimate marker-QTL linkage for many important yield traits that are measured on plots rather than on individual plants. For this purpose, it would be useful to utilize marker-QTL linkage analyses based on defined lines or families that can be replicated in large numbers as necessary for plot evaluations (for convenience, both lines and families will henceforth be referred to as "lines"). A further advantage is that by scoring the quantitative trait on replicated progenies of each line, the standard error of line means can be reduced relative to the standard deviation of individuals. Consequently, the required number of lines for given power can be less than the required number of F_2 individuals. Since marker genotype of a line can often be determined on the basis of a single individual, the use of replicated progenies should allow a corresponding reduction in the number of individuals scored for markers. This will, of course, be at the expense of a larger number of individuals scored for the quantitative traits.

Statistical aspects of replicated progenies were briefly considered by Beckmann and Soller (1986) and examined in some detail by Ellis (1986), Burr et al. (1988), and Cowen (1988). These analyses are further extended in the present study.

Contribution No. 2613-E of the Agricultural Research Organization, 1989 series

Theory

Considered here are four types of lines derived from the F_2 of a cross between two inbred lines: F_3 families, produced by selfing individual F_2 plants; F_4 families, each produced by bulk-selfing of numerous F_3 individuals derived from a single F_2 plant; vegetative clones; and recombinant inbred lines (RIL). A fifth type, doubled haploid lines, have the same statistical features as RIL, when the proportion of recombination between marker and QTL is zero and, hence, will not be specifically considered in this analysis. In the case of F_3 and F_4 families, only the F_2 parent of each family would be scored for the markers. In the case of vegetative clones or RIL, only a

single individual from each line would be scored. In all cases a number of individual from each of the progeny groups are evaluated for the quantitative traits.

Linkage between a marker-locus and a QTL in the F_2 of a cross between two inbred lines, or in the replicated F_2 derivatives, is detected as a significant difference in quantitative trait value between alternative homozygous genotypes at the marker locus (Soller et al. 1976). The number of F_2 offspring or of replicated lines per marker class (N) required for given power ($1-\beta$) and Type I error (α) will equal:

$$N = V(z_{\alpha/2} + z_\beta)^2/[2(1-2r)^2 d^2], \qquad (1)$$

where r is the proportion of recombination between marker locus and QTL, $2d$ is the difference in quantitative trait value between alternative homozygous genotypes at the QTL, V is the variance between the individual's means of all lines that share the same homozygous marker genotype, $z_{\alpha/2}$ and z_β are the ordinates of the standard normal curve correspondig to $\alpha/2$, and β, repectively.

As compared to the F_2 population or to one another, the various types of lines will differ in three factors:

(i) The proportion of all lines that fall into the informative (homozygous) marker classes. If T is the total number of lines scored for markers, then $N = T/4$ for vegetative clones, F_3 and F_4 families, and $N = T/2$ for RIL.

(ii) The proportion of recombination between marker and QTL. In particular, (Haldane and Waddington 1931) since the repeated selfing required in order to generate RIL allows additional rounds of recombination to take place as compared to the other types of lines, the proportion of recombination in RIL tends to $R = 2r/(1+2r)$ in the limit. This will not affect vegetative clones, F_3 and F_4 families since, in these cases, it is the F_2 linkage relationships that determine marker-associated quantitative effects due to linked QTL.

(iii) The variance (V) between the means of lines sharing the same homozygous marker genotype. This variance has two genetic components: V_A, due to recombination between marker and QTL (Ellis 1986) and V_B, due to segregation of loci not in linkage to the marker; and a residual component, V_W, due to random error. The effect of progeny replication on V can be derived as follows. Let,

$$Y_{ijk} = u + a_{ij} + b_{ij} + e_{ijk},$$

where Y_{ijk} = is the value of the kth replicate of the jth line having the ith marker genotype ($i = 1, 2$ depending on whether marker genotype is MM or mm; $j = 1, \ldots, N$ lines per marker genotype; and $k = 1, \ldots, n$ replicates per line. The replicates can represent individual progeny of the lines or individual plots each comprising many progeny); u is the overall mean.

a_{ij} is the expected genetic value of the jth line having the ith marker genotype, with respect to the QTL in linkage to the marker. The a_{ij} will have a multinomial distribution, taking on various values, according to whether genotype at the QTL in the original F_2 parent from which the line is derived was AA, Aa, or aa, with probabilities depending on: marker genotype, the proportion of recombination between marker and QTL, and the type of line involved. In particular, in the F_2 population itself among individuals having the MM genotype, the proportion of individuals or lines having expected values d, h, and $-d$ (corresponding to QTL genotypes AA, Aa, and aa, respectively will be $(1-r)^2$, $2r(1-r)$, and r^2, respectively. The same will hold for vegetative clones derived from this F_2 population. For F_3 and F_4 the proportions will remain the same, but the expected values of lines derived from F_2 individuals having AA, Aa, and aa genotypes will be, respectively, d, $h/2$, $-d$ for F_3 families, and d, $h/4$, and $-d$ for bulk F_4 families. Within any marker genotype, V_A, the variance of the a_{ij}, will depend on the main (d) and dominance (h) effects associated with the linked QTL and on the proportion of recombination (r) between marker and QTL. Depending on the proportion of recombination between marker and QTL, V_A will be only a fraction of the overall genetic variance attributable to the QTL in linkage to the marker.

b_{ij} is the net genetic value in the jth line of the ith marker genotype, with respect to all other segregating quantitative loci affecting the trait. The b_{ij} will be approximately normally distributed with mean zero and variance V_B, which is equal to the overall genetic variance between lines, V_G, less the variance associated with the QTL linked to the marker. It follows from the above that $V_A + V_B \leq V_G$.

e_{ijk} is a random error effect associated with the kth replicate of the jth line within the ith marker genotype. The e_{ijk} will be normally distributed with mean zero, and variance V_W.

Taking, as an approximation, $V_A + V_B = V_G$, the variance between means of individual replicated lines within marker genotypes (Y_{ij}) will be closely approximated by

$$V(Y_{ij}) = V_G + V_W/n,$$

where n is the degree of replication of the individual line.

Note that the genetic component of variance between means of replicated lines (V_G) is not reduced by increasing the degree of replication (n). This can be reduced only by increasing N, the number of lines having a given marker genotype. That is, the standard error associated with the mean of N lines having a given marker genotype ($Y_i..$), each line value itself being a mean value for n replicates, will be:

$$SE\,(Y_i..) = V_G/N + V_W/Nn.$$

The genetic variance between lines (V_G) and the variance within lines (V_W) for various classes of progenies

derived from and F_2 population are given in Mather and Jinks (1971) in terms of the additive (*D*) and dominance (*H*) portions of the genetic variance between lines and the environmental variance (*E*) within lines. These variances are summarized in Table 1 for the various progeny types considered here, with modifications and additions as necessary for cases not considered explicitly in Mather and Jinks (1971).

On the assumption of codominance at all QTL (including QTL not linked to the marker) F_2 genetic variance, as a proportion of total F_2 phenotypic variance, will equal the narrow-sense heritability (h^2) of the trait in the F_2 population, while F_2 environmental variance will equal $(1 - h^2)$. On this assumption, Table 2 shows the expected variance between replicated progenies sharing the same marker genotype for the various types of lines in terms of the heritability (h^2) and environmental $(1 - h^2)$ components of variance of the simple F_2 population. Note that the genetic variance between progenies for RIL is twice the F_2 genetic variance, while for F_3 and F_4 families, the variance within progenies includes a genetic component as well as an environmental component.

Table 2 also gives expressions for the relative number of lines required for given power, as compared to the number of F_2 individuals required for the same power, taking into account the proportion of informative lines, recombination between marker and QTL, and effect of replication on the variance between lines sharing the same marker genotype. Note that the expression for RIL also includes a factor involving the proportion of recombination between marker and QTL. This is due to the fact that, as noted above, recombination between marker locus and QTL for RIL will be equal to $R = 2r/(1+2r)$ rather than r and, hence, required numbers for RIL will be proportional to $1/(1-2R)^2$ rather than $1/(1-2r)^2$ (Haldane and Waddington 1931).

Consideration of Table 2 shows that except for RIL, the relative number of lines required for given power, as compared to the number of F_2 individuals required for the same power, tends to h^2 as n increases. For RIL, the relative number tends to $h^2[1-2r]^2/[1-4r/(1+2r)]^2$. The approach to the limit is rapid, being proportional to $1/n$ for vegetative clones, F_3 and F_4 families and to $1/2n$ for RIL. Thus, relatively little is to be gained by replication for traits having moderate to high heritabilities, and virtually all of the statistical benefits of replication will be obtained by $n = 10$. Benefits of RIL decrease rapidly with r. In fact for $r = 0.2$ and heritability of 0.5 or more, the utilization of RIL, even at maximum replication, will be less effective than a simple F_2 population. This can be avoided by increasing the number of markers so as to reduce the average spacing between marker and QTL, but this will, of course, require an increase in the number of markers required for marker-QTL mapping in RIL as compared to the other types of lines.

Table 1. Genetic variance between means of lines (V_G) and variance within lines (V_W) for various types of replicated progenies, in terms of additive (*D*) and dominance (*H*) portions of genetic variance between lines and environmental variance between replicates within lines

Type of progeny	V_G	V_W
F_2	$D/2+H/4$	E
F_3 [a]	$D/2+H/16$	$D/4+H/8+E$
F_4 [b]	$D/2+H/64$	$3D/8+3H/32+E$
RIL [c]	D	E

[a] $V_G = V_{1F3}$ and $V_W = V_{2F3}$ in Mather and Jinks (1971), not including environmental terms
[b] $V_G = V_{1F4}$ and $V_W = V_{2F4} + V_{3F4}$ in Mather and Jinks (1971), not including environmental terms
[c] Calculated according to procedure of Mather and Jinks (1971)

Table 2. Expected variance between and within lines, and number of replicates required for given power relative to the number of F_2 individuals required for the same power, for the various classes of replicated progenies. Values are given in terms of the heritability (h^2) and the environmentally determined (1-h^2) components of variance of a simple F_2 population and the degree of progeny replication (n) [a]

Progeny type	Variance component		Required no. of progenies relative to F_2 [b]
	Between progenies	Within progenies	
F_2	h^2	$(1-h^2)$	1.0
F_3	h^2	$(1-h^2/2)/n$	$h^2+(1-h^2/2)/n$
F_4	h^2	$(1-h^2/4)/n$	$h^2+(1-h^2/4)/n$
Vegetative clones	h^2	$(1-h^2)/n$	$h^2+(1-h^2)/n$
RIL	$2h^2$	$(1-h^2)/n$	$[h^2+(1-h^2)/2n]$ $[1-2r]^2/$ $[1-4r/(1+2r)]^2$
DHL [c]	$2h^2$	$(1-h^2)/n$	$h^2+(1-h^2)/2n$

[a] It is assumed that alleles at all QTL are codominant
[b] Taking into account that all RIL are informative for all markers, but for the other types of lines, only half of the lines will be informative for any particular marker
[c] Doubled haploid lines. Values are equal to those for RIL with proportion of recombination, $r = 0$

Discussion

The total number of replicated progenies required for statistical significance and good power in marker-QTL mapping can be quite large. Application of Eq. (1) to an F_2 population, assuming $2d/\sigma = 0.3$, $r = 0.05$, shows that 1,260 individuals would be required to provide Type I error 0.01 and power 0.80. Thus, depending on heritability, many hundreds of lines can be required for equivalent power. This may be feasible for vegetative clones, F_3, or F_4 families but does not seem practical for RIL or doubled haploids.

The decision as to whether to use replicated progenies for marker-QTL linkage studies and the type of replicat-

ed progeny to use will depend on many agrotechnical factors. Prominent among these will be the ultimate utilization of the information – as part of a breeding program, or as a basis for genetic studies of the quantitative trait. A discussion of these considerations is beyond the competence of the authors, and this study is consequently intended only to clarify the statistical aspects that can assist the breeder or geneticist in arriving at a decision. Nevertheless, it may be useful to point out that a major technical advantage of vegetative clones, doubled haploid lines, and RIL, independent of any effect of replication on required numbers, lies in the fact that the lines can be reproduced indefinitely and continually evaluated with respect to additional quantitative traits and markers, with all of the information cumulative (Burr et al. 1988).

Thus, in contrast to F_2 populations, which can be scored for quantitative traits at only one time and in only one place, once a set of vegetative clones, doubled haploid lines, or RIL have been throughly mapped with respect to markers, additional quantitative traits can be mapped by simply raising additional replicates of each line and evaluating these with respect to the quantitative trait alone; it will no longer be necessary to score the lines for markers. In this context, it should be noted that F_3 and, particularly, F_4 families share this advantage to a great extent. Although they cannot be replicated indefinitely, many seeds can be obtained from a single F_2 plant, and even more from an entire F_3 family. These can be stored for long periods. Since most of the advantage of replication is obtained with rather small replicates ($n \approx 10$), a set of F_3 or F_4 seeds could serve for evaluation of quantitative traits on many different occasions, even for traits measured on plots rather than individuals. Thus, from a purely statistical and technical point of view, F_3 and bulk F_4 families would appear to provide many of the advantages of doubled haploid lines or RIL for marker-QTL linkage studies, while being much easier to produce in large numbers.

Acknowledgements. The useful comments of Y. Kashi are acknowledged with thanks. This research was supported by a grant from the USA-Israel Binational Agricultural Research and Development Fund (BARD).

References

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms in plant genetic improvement. Oxford Surv Plant Mol Cell Biol 3:197–246

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Cowen NM (1988) The use of replicated progenies in marker-based mapping of QTLs. Theor Appl Genet 75:857–862

Edwards MD, Stuber CW, Wendel JF (1987) Molecular marker facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution, and types of gene action. Genetics 116:113–125

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

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

Kahler EL, Wehrhan CF (1986) Associations between quantitative traits and enzyme loci in the F_2 population of maize hybrids. Theor Appl Genet 72:15–26

Lebowitz RJ, Soller M, Beckmann JS (1986) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor Appl Genet 72:556–562

Mather KM, Jinks JL (1971) Biometrical genetics. Chapman and Hall, London

Nienhuis J, Helentjaris T, Slocum M, Ruggero B, Shaefer A (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop Sci 27:797–803

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

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

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

Weller JI, Soller M, Brody T (1988) Linkage analysis of quantitative traits in an interspecific cross of tomato (*Lycopersicon esculentum* × *Lycopersicon pimpinellifolium*) by means of genetic markers. Genetics 118:329–339