Theor Appl Genet (1988) 76:623–626

Detection of linkage between restriction fragment length polymorphism markers and quantitative traits

M. J. Asins and E. A. Carbonell
Instituto Valenciano de Investigaciones Agrarias, Apartado Oficial, E-46113 Moncada (Valencia), Spain

Received April 28, 1988; Accepted May 8, 1988
Communicated by H. F. Linskens

Summary. Methodologies commonly used to detect linkage of marker loci to loci affecting quantitative traits are discussed. It is shown that variances for the quantitative trait differ among marker genotypes when using F_2 or pooled backcross data if linkage exists. Hence, to analyze this type of data by single factor ANOVA or other statistical techniques that assume a common variance is inadequate. Restriction fragment length polymorphism (RFLP) markers are a powerful tool in plant breeding but cost is an important drawback; hence, a methodology is suggested to obtain the minimum number of plants in F_2 populations to detect such linkage.

Key words: Quantitative traits – Polymorphisms – Restriction fragments – Genetic markers

Introduction

Analysis of restriction fragment length polymorphisms (RFLPs) permits the rapid construction of detailed genetic maps (Botsein et al. 1980). They are fundamental tools for studies on selection, identification and organization of plant genomes (Tanksley 1983; Beckmann and Soller 1986; Landry and Michelmore 1987). RFLPs are being used in maize, tomato and lettuce to saturate or generate genetic maps (Helentjaris et al. 1986; Bernatzky and Tanksley 1986; Landry et al. 1987).

Methods for mapping, estimating costs and applications to plant breeding have also been reported (Beckmann and Soller 1983, 1986; Soller and Beckmann 1983). A general review on methods and applications of RFLP analysis to plants has been recently published (Landry and Michelmore 1987). As in other genetic markers like isozymes, an important application of RFLP is quantitative trait loci (QTL) investigations.

Statistical methods based on normal distribution (difference between means or single factor ANOVA) and three-point mapping are used to locate genome regions of loci contributing to quantitative traits (Gelderman 1975; Soller et al. 1976; Tanksley et al. 1982; Edwards et al. 1987).

The purpose of this paper is to show that if linkage is present, the genetic variance of the quantitative trait is not homogeneous among marker locus genotype classes in a F_2 population or in combined backcrosses; then, in some recently published papers, methodologies were wrong. Besides, once the genotypic variance formulae are deduced, the minimum number of offspring required to detect linkage of RFLP to QTL is obtained. Experimental plans to minimize that value are important since RFLP analysis costs are high.

Relation between marker and quantitative trait loci

Let us first consider F_2 populations derived from crosses between two inbred lines, denoting by P the marker locus and Q the locus involved in the expression of a quantitative character. Consider two codominant genes for the marker locus and that both loci are linked, r being the recombination fraction between them (2 r would be the probability of crossing over during meiosis) independent of the sex of the gametes produced. It is assumed that the environmental component of the variance of the quantitative trait does not depend upon the genotypes, i.e. it is randomly distributed.

For the genetic marker the segregation 1 P_1P_1:2 P_1P_2:1 P_2P_2 is expected in F_2. Defining a and d after Falconer (1960) as the genotypic value of the homozygote and the heterozygote genotypes, respectively, let us study the means and genotypic variances of the locus

Table 1. Within-class marker locus (P locus) frequencies for the quantitative trait (Q locus) in F_2

Genotypes Values	Q_1Q_1 a	Q_1Q_2 d	Q_2Q_2 −a
P_1P_1	$(1-r)^2$	$2r(1-r)$	r^2
P_1P_2	$r(1-r)$	$(1-2r+2r^2)$	$r(1-r)$
P_2P_2	r^2	$2r(1-r)$	$(1-r)^2$

Table 2. Within-class marker locus means and variances for the quantitative trait in F_2

Genotype	Mean	Variance[a]
P_1P_1	$a(1-2r)$ $+2dr(1-r)$	$2a^2r(1-r)+2d^2rs$ $-4adr(1-3r+2r^2)$
P_1P_2	$d(1-2r+2r^2)$	$2a^2r(1-r)+2d^2rs$
P_2P_2	$-a(1-2r)$ $+2dr(1-r)$	$2a^2r(1-r)+2d^2rs$ $+4adr(1-3r+2r^2)$

[a] $s=1-3r+4r^2-2r^3$

Table 3. Within-class marker locus means and variances for the quantitative trait in a backcross to line P_1P_1

Genotype	Mean	Variance
P_1P_1	$a(1-r)+dr$	$(a-d)^2r(1-r)$
P_1P_2	$ar+d(1-r)$	$(a-d)^2r(1-r)$

Table 4. Within-class marker locus means and variances for the quantitative trait in a backcross to line P_2P_2

Genotype	Mean	Variance
P_2P_2	$-a(1-r)+dr$	$(a+d)^2r(1-r)$
P_1P_2	$-ar+d(1-r)$	$(a+d)^2r(1-r)$

involved in the quantitative character within each genotypic class of the marker locus as a function of genotypic values, frequencies and the recombination fraction (Tables 1 and 2).

If both loci segregate independently, means and variances for the quantitative character are the same among all genotypes of the marker. When loci are syntenic, both means and variances are different among genotypes of the marker. Only when $d=0$ (no dominance, i.e. the genotypic value of the heterozygote is the average of both homozygotes) or $r=0$ (complete linkage), are variances the same. Furthermore, given that the expression $(1-3r+2r^2)$ is always positive in the interval $0\leq r\leq 0.5$, the P_1P_1 marker class should have the lowest variance and the P_2P_2 marker class the highest.

If backcrosses instead of F_2 populations are used, the situation changes. Tables 3 and 4 show the means and genotypic variances of the quantitative trait within the homozygote and heterozygote classes of the marker in backcross to line P_1P_1 and to line P_2P_2, respectively. In a given backcross progeny, means are different between genotypes at the marker locus when linkage exists but variances are the same. However, variances are also different when data from backcrosses to both parental lines are combined.

Denoting by N the number of individuals screened per homozygote, M_1 and M_2, and V_1 and V_2 as the means and variances of the quantitative trait for the P_1P_1 and P_2P_2 genotypic classes of the marker, then if linkage exists the following condition must be satisfied in order to be statistically detectable at the 5% level of significance with a probability of error Type II of 0.10 ($z_{\alpha/2}=1.96$ and $z_\beta=1.28$):

$$3.24\leq(M_1-M_2)/\sqrt{(V_1+V_2)/N}$$

where

$$M_1-M_2=2a(1-2r)$$

and, assuming initially for the clarity of the presentation of results that the environmental variance is negligible,

$$V_1+V_2=4a^2r(1-r)+4d^2r(1-3r+4r^2-2r^3). \quad (1)$$

Then, if $d=ka$ for some constant value k depending on the degree of dominance of the quantitative trait ($k=0$ implies no dominance; $k=1$ complete dominance; and $k>1$ overdominance), the above condition can be rewritten as:

$$N\geq 10.5(r(1-r)+k^2r(1-3r+4r^2-2r^3))/(1-2r)^2. \quad (2)$$

This equation can be evaluated for given values of r and k, and so the minimum value of N that satisfies the condition is obtained. This relationship is shown in Fig. 1 for $k=0$, ¼, ½, ¾, 1 and 1.25. There are, of course, other contrasts involving heterozygotes that could be used in a similar fashion to detect linkage.

In order to have a more realistic approach, the amount of the environmental variance in the population as a whole should be introduced in the formula. Using the definition of heritability *sensu lato* h^2 as the ratio of the genotypic variance V_G to the total variance (genotypic variance plus environmental variance V_E) and assuming no genotype-environment interaction it can be shown that

$$V_E=(\tfrac{1}{2}a^2+\tfrac{1}{4}d^2)(1/h^2-1).$$

If again $d=ka$ and assuming that V_E does not change with marker genotype classes, which is quite reasonable, then (1) has the following extra term:

$$a^2(1+\tfrac{1}{2}k^2)(1/h^2-1)$$

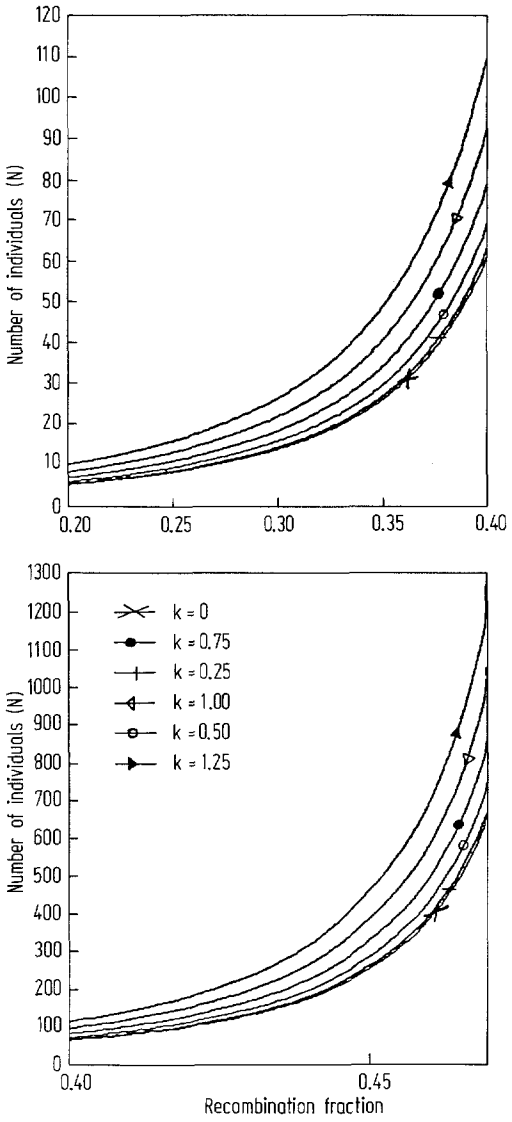

Fig. 1. Minimum sample size required to statistically detect linkage at the 5% level and probability of Type II error of 0.1, as a function of recombination fraction and degree of dominance assuming negligible environmental variance

and (2) becomes

$$N \geq 10.5[r(1-r)+k^2r(1-3r+4r^2-2r^3)+\tfrac{1}{4}(1+\tfrac{1}{2}k^2)\cdot(1/h^2-1)]/(1-2r)^2.$$

Therefore, the increase in sample size per homozygote in this case is:

$$2.625(1+\tfrac{1}{2}k^2)(1/h^2-1)/(1-2r)^2$$

for values

$0 \leq k \leq 1.25$, then $1 \leq (1+\tfrac{1}{2}k^2) \leq 1.8$
$0 \leq r \leq \tfrac{1}{3}$, then $1 \leq 1/(1-2r)^2 \leq 9$ and
$0.5 \geq h^2 \geq 0.1$, then $1 \leq (1/h^2-1) \leq 9$.

Hence, for the cases considered here, the worst combination of parameters implies 383 additional individuals per homozygote.

Discussion

This analysis shows that linkage can be theoretically detected between RFLP markers and genes involved in quantitative character variation, depending only on the number of individuals to be screened, which in turn is a function of the degree of dominance (k), the strength of linkage (r) and the heritability of the quantitative trait (h^2). We have shown that linkage makes genotypic variances for the quantitative character differ among genotypic marker classes in F_2 populations or in combined backcrosses. Hence, there is a widespread mistake in the bibliography when considering the genetic variance within marker class as having a common value. Therefore, with experimental data, if there is no previous knowledge of the type of gene action, the first step would be to test the homogeneity of variances. If it is significant it could be an indication of linkage. In this case, to detect linkage the contrasts between means should be performed using the estimates of the variances concerning the genotypes whose means are compared and not the error term of the analysis of variance. If the homogeneity test is not significant and the means are significatively different, it could be an indication that there is no dominance for the quantitative character in the F_2 population and that linkage is present. In this case, the analysis of variance prior to the mean contrasts would be adequate to detect linkage.

It is important, before using RFLPs as genetic markers and depending on the sample size to be screened, to study the cost involved in the experiment (Beckmann and Soller 1983). The method presented here is a realistic approach to the problem. Ellis (1986), using recombinant inbred lines, showed a relationship between the minimum recombination frequency for which linkage to an RFLP marker can be detected when parents differ by a given number of standard deviation units and a different number of recombinant lines are used. This author assumes that the environmental component of the variance is negligible. Another commonly cited reference regarding our conclusions is that of Soller et al. (1976), but they assumed complete linkage ($r=0$) of the marker locus to the quantitative loci in their initial approach (the genotypic variance within marker locus classes will be zero in this case). They calculated the number of offpring required to detect a quantitative locus whose effect is that of a locus contributing an amount of additive genetic variance equal to 1% of the total phenotypic variance in the F_2 generation. When they considered the effect of recombination they simply suggested that sample sizes for a given power would have to be increased by $1/(1-2r)^2$ in a

normal-test situation, and that for the analysis of variance with r and d not equal to zero, the power of the experiment is reduced even further. In their calculations they considered the within class marker variance (their σ^2_w) to be homogeneous among marker classes in F_2 and combined backcross data for the whole interval of existence of r ($0 \leq r \leq 0.5$), which we have shown is wrong unless the environmental variance buffers the genotypic variance differences among marker locus genotypes. Quite interestingly, Edwards et al. (1987) have found some homogeneity tests of variance among marker locus genotypes to be significant when studying the linkage between isozyme loci to QTL. The interpretation of these findings seems to be as easy, as it should be the rule and not the exception for any value of *r* and *d* different from zero.

References

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

Beckmann JS, Soller M (1986) Restriction length fragment polymorphisms and genetic improvement of agricultural species. Euphytica 35:111–124

Bernartzky R, Tanksely SD (1986) Towards a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887–898

Botsein D, White RL, 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

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

Falconer DS (1960) Introduction to Quantitative Genetics. Ronald Press, New York

Gelderman H (1975) Investigations on inheritance of quantitative characters in animals by gene markers. 1. Methods. Theor Appl Genet 46:319–330

Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of linkage maps in maize and tomato using restriction length polymorphisms. Theor Appl Genet 72:761–769

Landry BS, Michelmore RW (1987) Methods and applications of restriction fragment length polymorphism analysis to plants. In: Bruening G, Harada J, Kosuge T, Holleander A (eds) Tailoring genes for crop improvement. Plenum, New York

Landry BS, Kesseli R, Hei Leung, Michelmore RW (1987) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (*Latuca sativa* L.). Theor Appl Genet 74:646–653

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

Soller M, Brody T, Genizi MA (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

Tanksley SD (1983) Gene mapping. In: Tanksley SD, Norton T (eds) Isozymes in plant genetics and breeding, part A. Elsevier, Amsterdam

Tanksley SD, Medina-Filho H, Rick CM (1982) Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49:11–25

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

Note added in proof

Since submission of this manuscript Weller et al. (1988) have found what they considered an unexpected large number of statistically significant marker-associated effects on variance. The comments throughout this manuscript and mainly that at the end of the Discussion also apply to that paper.