Theor Appl Genet (1994) 89:351–357

A. Darvasi · M. Soller

Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci

Received: 12 July 1993 / Accepted: 21 December 1993

Abstract The cost of experiments aimed at determining linkage between marker loci and quantitative trait loci (QTL) was investigated as a function of marker spacing and number of individuals scored. It was found that for a variety of experimental designs, fairly wide marker spacings (ca. 50 cM) are optimum or close to optimum for initial studies of marker-QTL linkage, in the sense of minimizing overall cost of the experiment. Thus, even when large numbers of more or less evenly spaced markers are available, it will not always be cost effective to make full utilization of this capacity. This is particularly true when costs of rearing and trait evaluation per individual scored are low, as when marker data are obtained on individuals raised and evaluated for quantitative traits as part of existing programs. When costs of rearing and trait evaluation per individual scored are high, however, as in human family data collection carried out primarily for subsequent marker – QTL analyses, or when plants or animals are raised specifically for purposes of marker – QTL linkage experiments, optimum spacing may be rather narrow. It is noteworthy that when marginal costs of additional markers or individuals are constant, total resources allocated to a given experiment will determine total number of individuals sampled, but not the optimal marker spacing.

Key words Quantitative trait locus · Genetic mapping · Marker-QTL linkage · Experimental design

Introduction

With the advent of DNA-level genetic markers (Beckmann and Soller 1983, 1990; Botstein et al. 1980) it is possible to cover the genome completely with a more or less evenly spaced set of segregating genetic markers in specific crosses or families. In a single experiment, such a spanning set allows linkage relationships between marker loci and quantitative trait loci (QTL) that affect a particular trait or group of traits to be evaluated for an entire genome in a single experiment. A variety of experimental designs provide data amenable to such analyses, including F_2 and backcross (BC) populations derived from crosses between inbred lines (Edwards et al. 1987; Darvasi and Soller 1992; Darvasi et al. 1993; Haley and Knott 1992; Kahler and Wehrhahn 1986; Knapp et al. 1990; Lander and Botstein 1989; Soller et al. 1976; Van Ooijen 1992; Weller 1987; Weller et al. 1988), replicated progenies and recombinant inbred lines (RIL) (Soller and Beckmann 1990), half-sib families (Beever et al. 1990;.Geldermann et al. 1985; Gonyon et al. 1987; Soller and Genizi 1978; Weller et al. 1990), and relative-pairs (Amos and Elston 1989; Blackwelder and Elston 1985; Cockerham and Weir 1983; Hill 1975; Penrose 1938).

Within any particular experimental design, the power of marker – QTL linkage studies depends on the number of individuals raised and scored for markers and quantitative traits, on the map distance between markers and QTL, on the phenotypic effect of the QTL, and on the error variance and the designated type I error (Amos and Elston 1989; Darvasi et al. 1993; Lander and Botstein 1989; Soller et al. 1976; Soller and Genizi 1978).

Classically, the number of experimental individuals is a parameter that is subject to control. As progressively larger numbers of well-mapped markers become available, however, the average spacing between markers also becomes subject to experimental control. It is of interest, therefore, to examine the total cost for marker-QTL linkage studies as a function of both marker spacing and the number of individuals scored, in order to determine the optimum experimental structure in the sense of minimizing total costs for given power. The experiments analyzed here are those intended to provide initial information as to the general distribution of QTL

Communicated by A. L. Kahler

A. Darvasi (✉) · M. Soller
Department of Genetics, The Alexander Silberman Life Sciences Institute, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

affecting a particular trait across the entire genome. Once individual QTL or QTL clusters have been assigned to specific chromosomal regions, additional experiments may be necessary to map them more exactly with these regions. Optimum experimental designs for such second-stage experiments are not considered here.

Theory

The cost of a marker-QTL mapping experiment, as a function of marker spacing, M, has the form:

$$f(M) = c_1 n(M) b(M) + c_2 n(M) \quad (1)$$

where, M = map distance between adjacent markers, in Morgans (1 Morgan = 100 cM); $n(M)$ = the number of individuals required for marker-QTL linkage determination with given power and type I error, as a function of M, where power is averaged over all possible locations of a QTL in the interval M; $b(M)$ = the number of markers scored per individual, as a function of M. It is immediately evident that $b(M) = G/M$, where G = total genome size in Morgans; c_1 = genotyping costs of a single marker; c_2 = phenotyping costs of a single individual, including: rearing, trait evaluation and DNA sampling.

Optimum marker spacing is obtained by minimizing $f(M)$ as a function of marker spacing, M. In some cases, the costs c_1 and c_2 will not remain constant as the number of individuals or markers increases. These cases will be considered in detail in a later section.

Minimizing $f(M)$ with c_1 and c_2 constant

In many cases, costs per individual and costs per marker will be essentially independent of the number of individuals sampled and the number of markers scored. In these cases c_1 and c_2 will be constant and dividing Eq. 1 by c_2 reduces the cost function to the form:

$$f^*(M) = n(M)(Gc + M)/M \quad (2)$$

where c is the ratio c_1/c_2 and $f^*(M)$ is a proportional cost function with the same optimum as $f(m)$.

Expressions for $n(M)$, the number of individuals required for marker-QTL linkage determination with given power and type I error rate, as a function of marker spacing, M, are not provided directly in the literature. Rather, expressions for required sample size are given as a function of the proportion of recombination, r, between markers and QTL. Such expressions are given in the literature for analyses involving single markers in BC and F_2 designs (Soller et al. 1976), replicated progenies and RIL designs (Soller and Beckmann 1990; Knapp 1991) and half-sib families (Weller et al. 1990). For the purposes of the present analysis, these expressions can be considered to consist of two components: one, denoted $h(r)$, a function of r; the other, denoted k, a function of all of the other parameters determining required sample size for a given experimental design, i.e., QTL effect, error variance, and type I and type II error rate.

For all of the above experimental designs (expect RIL and relative pairs) $h(r)$ was found to have the form $h(r) = 1/(1-2r)^2$. Expressions for RIL were obtained by substituting $r^* = 2r/(1+2r)$ for r (Haldane and Waddington 1931). Expressions for relative pairs were taken as numerical functions obtained from Table IV of Amos and Elston (1989).

Expressions for $h(r)$ were transformed to the corresponding map-distance expressions, denoted $h(m)$, (where m is the map distance in Morgans between the QTL and the nearest marker), by means of the Haldane mapping function, $r = (1 - e^{-2m})/2$ (Haldane 1919). $n(M)$ was then obtained by averaging $kh(m)$ over all possible locations of the QTL, relative to the nearest marker (i.e., over the interval 0 to M/2). Averaging was carried out by assuming that a QTL is equally likely to be found at any point within the region M, i.e., that m, the distance between QTL and the nearest marker, has a uniform distribution over the interval 0 to M/2, thus a uniform density function, $g(m)$,

$$g(m) = 1/(M/2 - 0) = 2/M \quad (3)$$

Thus, for these designs and analyses, $n(M)$ can be written as:

$$n(M) = 2k/M \int_0^{M/2} h(m)\, dm \quad (4)$$

Relationship of marker spacing and all other parameters (k)

Although M does not appear directly in the parameters of k, k does include type I and type II errors, both of which are affected by M, albeit in opposite directions. Namely, as M decreases, more markers are included, and type I error should be adjusted downwards to obtain the overall type I error desired. At the same time, when more markers are included, type II error will decrease, since with more markers in the vicinity of any given QTL there is a higher likelihood that one or another of the markers will give a significant result. Consequently, type II error for any individual marker test should be adjusted upwards in order to obtain a given overall type II error. Since in k, type I and type II errors appear as the sum of their standard normal distribution ordinates $z_{\alpha/2}$ and z_β (see following example), this reduces the overall effect on k. It should also be noted that when M is small, further decrease in M has no further effect on type I and type II errors, while for large M the effects of small changes in M on type I and type II errors are mild (Lander and Botstein 1989). Thus, if all factors are taken into consideration it would appear that within the range of values for M over which

optimization takes place, net variation in k as a result of M acting through type I and type II errors will be negligible. Consequently, optimum spacing, M, which minimizes the cost function, will not be affected by k and can be analyzed as a function of $h(r)$ only.

Illustrative example

In order to illustrate the above procedure, the expression for the cost function (Eq. 2) for an F_2 analysis carried out by ANOVA methods will be derived explicitly. This will require deriving $n(M)$ as given in expression 4. Following Soller et al. (1976) for the F_2 case, the number of individuals, N, needed to obtain a given power equals

$$N = 8\sigma^2(z_{\alpha/2} + z_\beta)^2/[d^2(1-2r)^2] \quad (5)$$

where, σ^2 is the population variance, d is the QTL gene substitution effect (equal to half the difference between the expected quantitative value of the two homozygous QTL genotypes), and α and β are designated type I and type II error rates, respectively. Using the above notation, the components $h(r)$ and k of N are

$$h(r) = 1/(1-2r)^2, \quad (6)$$

and

$$k = 8\sigma^2(z_{\alpha/2} + z_\beta)^2/d^2. \quad (7)$$

$h(m)$ can now be obtained from $h(r)$ by the substitution, $r = (1-e^{-2m})/2$ according to the Haldane mapping function to give:

$$h(m) = 1/[1-2(1-e^{-2m})/2]^2 = 1/e^{-4m} = e^{4m} \quad (8)$$

Substituting $h(m)$ in Eq. 4 gives

$$n(M) = 2k/M \int_0^{M/2} (e^{4m})\,dm = k(e^{2M}-1)/2M \quad (9)$$

Substituting this expression for $n(M)$ in Eq. 2 gives

$$f^*(M) = k(e^{2M}-1)(Gc+M)/2M^2 \quad (10)$$

Since an analytical solution for the optimization of $f^*(M)$ was not available, a numerical optimization was carried out. The optimum found for this example would be the same for all cases where $h(r) = 1/(1-2r)^2$.

The genotyping and phenotyping costs c_1 and c_2

The analysis carried out assumes constant values for c_1 and c_2, as previously defined. This assumption will be correct in many experiments. For example, in determining linkage between marker loci and QTL in dairy cattle using the "Daughter" or "Granddaughter" designs (Weller et al. 1990), trait information is available on many more animals than will eventually be sampled, and the costs of obtaining the first blood or semen sample are essentially the same as those of obtaining the n^{th} sample. The same will apply for plant genetic programs when the total number of plants reared for mapping purposes is well within the existing overall resources of the organization involved. In this case, sample size will affect allocation of resources within the organization, but will not require specific investment in additional resources, so that costs per plant will be independent of the number of plants or plots dedicated to the QTL-marker linkage study. Similarly, for some of the newer classes of genetic marker uncovered by polymerase chain reaction (PCR) techniques (Beckmann and Soller 1990), costs per marker are independent of the number of markers scored per individual.

Although a constant value for c_1 and c_2 is often a reasonable approximation, in some instances it will clearly not apply. For example, there might be at least four stages of data collection in human genetics, with increased c_2 costs at each stage: individuals ascertained through a clinic, relatives within the geographical area of the study, and relatives in distant locations. Finally, at some point the available pedigrees ascertained through the local clinic may be exhausted; additional pedigree collection may then require very expensive methods. For animals that require physical space, increasing the sample size beyond some limit may involve building or buying new facilities. For many plants, an increase in sample size might require acquiring additional field space or constructing additional greenhouses, etc. Similarly, substrates and stains for isozyme markers, and restriction enzymes for restriction fragment length polymorphism (RFLP) markers, vary widely in costs. Hence, the c_1 costs of additional markers scored per individual can also increase with increased numbers of markers. Marker costs are also related to marker information content (Botstein et al. 1980). That is, for any particular marker some greater or lesser proportion of parents and offspring (depending on the number and frequency of alleles at the marker) will not contribute information useful for linkage analysis. Optimal experimental design will require approximately equal degrees of informativeness for the various markers used in the analysis. For some chromosomal regions, this will require utilizing more than one marker. In this case, it would seem appropriate to consider the effective cost of the marker for this region as the total cost of all of the markers used to obtain the desired degree of informativeness.

Minimizing $f(M)$ with changing costs of c_1 and c_2

From the above examples, it can be seen that the effect of increasing costs with increasing number of individuals sampled will often be stepwise in nature. In this case,

when sampling N individuals, there will be an initial cost, denoted I_1, independent of N, and then some constant cost, denoted $c_{2,1}$, for each individual from $N=0$ to $N=n_1$, where n_1 is the maximum number of individuals that can be sampled at cost $c_{2,1}$. Above this number, there would be a further initial cost, I_2, independent of N, and then some further constant cost, denoted $c_{2,2}$, for each individual from $N=n_1+1$ to $N=n_2$, where n_2 is the maximum number of individuals that can be sampled at a cost of $c_{2,2}$ or less. Additional cost levels for further increase in total sample size beyond n_2 can be denoted similarly. Thus, in general, the cost of sampling x individuals, denoted as a function $c_2(x)$ would be:

$$c_2(x)=\begin{cases} I_1+c_{2,1}\,x & 0<x\le n_1 \\ I_1+c_{2,1}\,n_1+I_2+c_{2,2}(x-n_1) & n_1<x\le n_2 \\ I_1+c_{2,1}\,n_1+I_2+c_{2,2}(n_2-n_1) & \\ \quad +I_3+c_{2,3}(x-n_2) & n_2<x\le n_3 \\ \vdots & \end{cases}$$

Since N will be related to anticipated gene effect, error variance and type I and type II errors, it will be possible to obtain an a priori initial estimate of the sample size range, n_{k-1} to n_k, in which N lies. For this estimate,

$$c_2(N)=I_1+c_{2,1}n_1+I_2+c_{2,2}(n_2-n_1)$$
$$+\ldots+I_k+c_{2,k}(N-n_{k-1})$$

or,

$$c_2(N)=I+c_{2,k}(N-n_{k-1}) \quad n_{k-1}<N\le n_1$$

where, I represents all of the constant terms. Since I is a constant, it will not affect the optimum. Consequently, using $c_{2,k}$ the marginal cost of sampling an individual in the range $n_{k-1}+1<N<n_k$ in Eq. 1 will provide optimum marker spacing. A similar argument can be made for cases where a number of cost levels are present for the markers scored. Thus, this derivation shows that for the non-linear "stepped" case, the appropriate analysis for minimizing $f(M)$ is as previously described, except for c_2 and c_1, which will receive the values of the marginal costs for the last individuals sampled and the last or most costly markers scored, respectively.

When differences in marginal cost levels are small (either for individuals or for markers) it will not be possible to obtain *a priori* estimates of the range in which the optimum number of individuals or markers is found. Consequently, the cost of the last individual sampled and the last marker genotyped will not be known. In this case, an iterative solution can be implemented, where in each step the estimated range of N is adjusted according to the estimate given by the prior step until the total cost agrees with the values obtained for "optimum" sample size and number of markers.

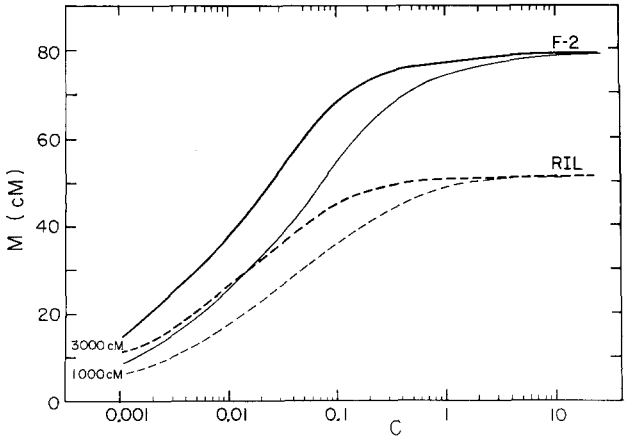
Numerical results

Figure 1 shows optimum marker spacing as a function of the ratio (c) between the cost of scoring a single marker and the cost of raising a single individual for an F_2 design and for an RIL design. The results for an F_2 design are also valid for single-marker analyses carried out by ANOVA in backcross and half-sib family designs. Optimum spacing was investigated for two genome sizes: 1000 cM, representing a typical plant genome, and 3000 cM, representing a typical animal or polyploid plant genome. Optimum spacing was always greater for a total genome size of 3000 cM than for a genome size of 1000 cM. The difference was relatively slight when the costs of marker determinations were high relative to the costs of raising individuals ($c>0.5$), but it became marked (spacings up to twofold greater for the larger genome size) when the costs of marker determinations were much less than the costs of raising individuals ($c<0.005$). This is plausible, since at any given spacing, the effect of greater genome size is to increase the relative weight of the many more marker determinations required per individual, relative to the costs of raising an individual.

Optimum spacing for RIL experiments was about two-thirds that of the F_2 experiments over the entire range of G and c values. This is a direct result of the fact that for given marker spacing, the proportion of recombination between any given marker and nearest QTL will be much greater for a population of RIL than for an F_2 population.

Optimum spacings for grandparent-grandchild pairs were virtually identical to those obtained for BC and F_2 populations (Fig. 2). Spacings for sib and half-sib pairs were very similar to one another but considerably narrower over all values of c than spacings for grandparent-grandchild pairs. Spacings for avuncular (i.e., uncle and

Fig. 1 Optimum marker spacing (M) for F_2 (equal to backcross data or half-sib families) and for recombinant inbred lines (*RIL*) as a function of genome size (G) and relative cost (c, shown in logarithmic scale) of scoring a single marker in a single individual relative to cost of rearing, trait evaluation, and DNA sampling for a single individual



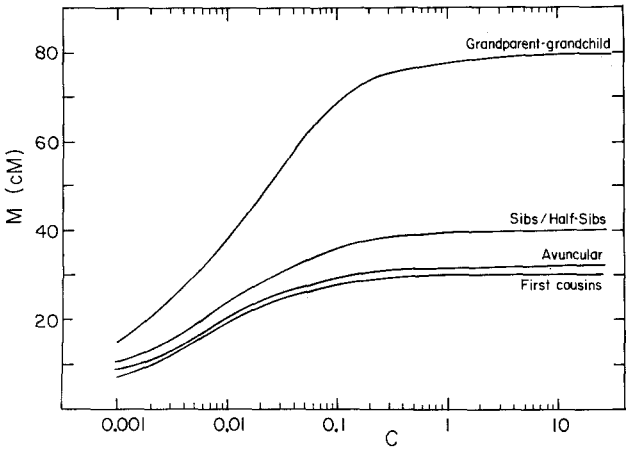


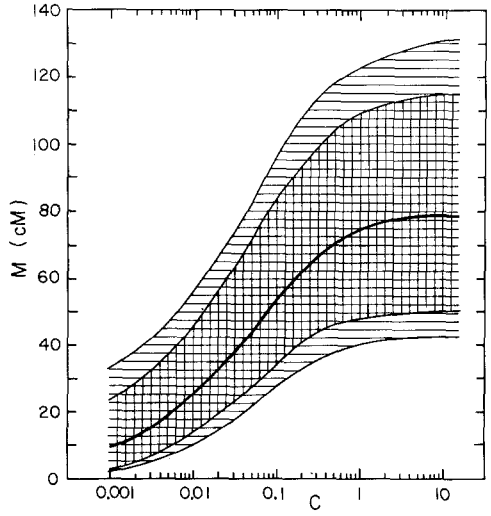
Fig. 2 Optimum marker spacing (M) for relative-pairs. Spacing is for relative pair analysis of pedigree data, based on single markers, as a function of relative cost (c, shown in logarithmic scale) of scoring a single marker in a single individual relative to cost of rearing, trait evaluation, and DNA sampling for a single individual

aunt to nephew or niece) and cousin pairs were narrower yet, and similar to one another. An intuitive explanation for these effects is not evident.

In order to examine the effect of deviation from optimum spacing on total experimental costs for given power, spacing contours giving total costs within 10% or 20% of the optimum were obtained separately for F_2 experiments, for RIL analyses and for genome sizes of 1000 cM and 3000 cM. It was observed that the contour widths for a given optimum spacing were virtually identical for the different analyses except for somewhat narrower widths for RIL in the range $c \leq 0.1$ (data not shown). Hence, although only curves showing contours for the F_2 1000 cM genome size situation are presented (Fig. 3), these can be taken to be representative contours for the other analyses as well. The range of marker spacings within 10% of the optimum was rather large in all cases, indicating that within this range deviation from optimum spacing does not greatly affect costs for given power. The range for marker spacing at 20% of the optimum was only slightly wider, indicating that costs for given power increase steeply beyond a 10% deviation from the optimum.

For BC and F_2 designs, a marker spacing of 30 cM is generally within 10% and always within 20% of the optimum for $c \leq 0.1$. This also holds for grandparent-grandchild pairs. For RIL and the remaining types of relative-pairs, the corresponding value was 20 cM. For all designs, a marker spacing of 45 cM is generally within 10% and always within 20% of the optimum for $c \geq 0.05$.

Fig. 3 Marker spacings providing total costs within 10% (*double hatching*) and 20% (*single hatching*) of the minimum. Values are presented for F_2 backcross or half-sib families, with G = 1000 cM, as a function of relative cost (c, shown in logarithmic scale) of scoring a single marker in a single individual relative to cost of rearing, trait evaluation, and DNA sampling for a single individual



Discussion

The major result of these analyses is the finding that for a wide range of experimental designs, fairly wide marker spacings are optimum or close to optimum for the initial studies of marker-QTL linkage in the sense of minimizing overall costs of the experiments. This holds true for experiments involving: (1) crosses between inbred lines or between lines within selfing species, typical of studies in many agricultural plants; (2) half-sib families, typical of studies in agricultural animals; and (3) relative-pairs, typical of studies in human genetics. The result also holds true over a wide range of relative costs of scoring a single marker in a single individual, as compared to the costs of rearing, evaluation and sampling ($c > 0.1$). Although the very wide marker spacings that are optimal in these cases may seem counter-intuitive, it should be realized that even for a marker spacing of 70 cM, the maximum distance between a QTL and nearest marker will be 35 cM, equivalent to $r = 0.25$; while the average distance between a QTL and the nearest marker will be only 17.5 cM, equivalent to $r = 0.15$. Similarly, for a marker spacing of 35 cM, the average distance between a QTL and the nearest marker will be equivalent to $r = 0.09$. Thus, very large spacings between markers in cM translate into a relatively low average proportion of recombination between marker and nearest QTL.

From the assumption of constant k, it follows that sample size will determine power for a given type I error, gene effect and error variance, but for a given experimental design, the same optimal marker spacing will provide maximum power for any combination of these factors. It also follows that when marginal costs of additional markers or individuals are constant, total resources allocated to a given mapping experiment will affect the number of individuals sampled, but not the optimal marker spacing. This will be the situation for the newer classes of DNA-level markers involving PCR and for many situations involving agricultural or experimental species. This means that in such cases optimal

marker spacing for maximum overall power is unaffected by budgetary considerations. That is, if budgets are reduced, maximum power will be obtained if the entire budget reduction is met by reducing the number of individuals utilized in the experiment. Conversely, if budgets are increased, power will be increased most effectively by increasing the number of individuals at the same marker spacing. This will not be the case when marginal costs of additional markers or individuals vary. In this case, a change in experiment size can change the marginal costs ratio, and hence the optimal marker spacing.

When data are analyzed using likelihood-ratio tests and interval mapping (Darvasi et al. 1993; Jensen 1989; Knapp et al. 1990; Lander and Botstein 1989), the number of offspring required for given power increases less markedly with r than when ANOVA and single-marker analysis are used. This effect, however, was found to be slight (Darvasi et al. 1993). Nevertheless, when likelihood ratio tests are used, optimal marker spacings will be slightly wider than those presented here.

Relative costs in the range $c \geq 0.1$ might be obtained when data and DNA samples are obtained from individuals that are otherwise reared and evaluated for other purposes (e.g., in dairy cattle recording and progeny testing schemes, existing fruit tree plantations, or ongoing family studies for behavioral or health traits in man). In this case, costs include only data retrieval (generally negligible) and DNA sampling, which can be somewhat expensive ($25–$50) per sample, depending on the dispersal of individuals and the costs of extracting DNA from tissue. When plants or animals are raised specifically for purposes of marker-QTL linkage experiments, however, or when human family data collection is designed specifically for subsequent marker-QTL analyses, relative costs can be in the range $c = 0.01$, for annual plant, poultry, or mouse experiments, to $c = 0.001$, for fruit trees, larger livestock, or man. The parameter c will, of course, also depend on the cost of evaluating the traits under analysis. For example, in experiments currently underway in Africa, that are aimed at mapping loci conferring trypanotolerance through the analysis of crosses between trypanotolerant N'Dama and trypanosensitive Zebu cattle (Kemp 1992; Soller 1992), F_1, F_2 and BC animals are produced specifically for this purpose and must be individually evaluated for trypanotolerance and associated traits. In this case, cost per animal may be more than a thousandfold greater than the cost of an individual marker evaluation, and close marker spacing will be optimal.

Trial calculations were carried out in order to provide some notion of the sample sizes and number of markers scored per individual that will be required in order to obtain given power for a typical marker-QTL experiment at optimum marker spacing. The case investigated was a BC design with ANOVA analysis, assuming co-dominance at the QTL, and expected mean difference between alternative homozygous genotypes at the QTL equal to 0.5 standard deviation units. Calculations were carried out to determine the total sample sizes and the total number of markers scored that would be required in order to detect a QTL with a power of 0.80 and Type I error of 0.05 at optimum spacing. When $c \geq 0.1$ (i.e., it costs no more than ten times as much to rear and score an individual for the quantitative traits and obtain a DNA sample as it does to score an individual marker), a marker-spacing of about 60 cM was optimal, i.e., 17 markers per individual for a genome size of 1000 cM, 50 markers per individual for a genome size of 3000 cM. This marker spacing requires scoring about 1000 individuals for the desired power, i.e., 17 000 markers scored for $G = 1000$, 50 000 markers scored for $G = 3000$. When the costs of rearing, trait scoring and DNA sampling were 100 times that of scoring an individual marker ($c = 0.01$), optimal spacing decreased by one-half, to 30 cM, and the number of individuals by one-third, to about 700. The total number of markers scored is now 23 000 (35% greater) for $G = 1000$, and 70 000 (40% greater) for $G = 3000$. Only when the costs of marker scoring drop to one-thousandth of the cost of rearing, trait scoring and DNA sampling is it optimal to use really large numbers of markers for a spacing of say, 10 cM. Even in this case, it is necessary to raise about 600 offspring for the desired power. Thus, even with the full utilization of marker capacity, marker-QTL experiments will still require hundreds of individuals for good power against QTL of relatively strong effects.

Aknowledgements We thank Margaret Mackinnon and Joel I. Weller for useful comments. This study was supported by the U. S. -Israel Binational Agricultural Research and Development Fund (BARD) and the U. S.-Israel CDR program.

References

Amos CI, Elston RC (1989) Robust methods for the detection of genetic linkage for quantitative data from pedigrees. Genet Epidemiol 6:349–360

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

Beckmann JS, Soller M (1990) Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/Technology 8:930–933

Beever JS, George PD, Fernando RL, Stormont CJ, Lewin HA (1990) Association between genetic markers and growth and carcass traits in a paternal half-sib family of Angus cattle. J Anim Sci 68:337–344

Blackwelder WC, Elston RC (1985) A comparison of Sib-Pair linkage tests for diseases susceptibility loci. Genet Epidemiol 2:85–97

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

Cockerham CC, Weir BS (1983) Linkage between a marker locus and a quantitative trait of sibs. Am J Hum Genet 35:263–273

Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theor Appl Genet 85:353–359

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker facilitated investigation of quantitative trait loci in maize. I.

Numbers, genomic distribution and types of gene action. Genetics 116:113–125
Geldermann H, Peiper U, Roth B (1985) Effects of marked chromosome section on milk performance in cattle. Theor Appl Genet 70:138–146
Gonyon DS, Mather RE, Hines HC, Haenlein GFW, Arave CW, Gaun SN (1987) Association of bovine blood and milk polymorphisms with lactation traits: Holsteins. J Dairy Sci 70:2599–2609
Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309
Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374
Haley CS, Knott SA (1992) A simple regression method for mapping quantitative loci in line crosses using flanking markers. Heredity 69:315–324
Hill AP (1975) Quantitative Linkage: a statistical procedure for its detection and estimation. Ann Hum Genet 38:439–449
Jensen J (1989) Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker loci. Theor Appl Genet 78:613–618
Kahler AL, Wehrhahn CF (1986) Associations between quantitative traits and enzyme loci in the F_2 population of a maize hybrid. Theor Appl Genet 72:15–26
Kemp S (1992) Report on ILRAD work on the genetics of Trypanotolerance. In: Proc FAO Expert Consultation Genet Aspects Trypanotolerance. FAO, Rome pp 122–125
Knapp SJ (1991) Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progeny. Theor Appl Genet 81:333–338
Knapp SJ, Bridges WC, Birkes D (1990) Mapping quantitative trait loci using molecular marker linkage map. Theor Appl Genet 79:583–592
Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199
Penrose GS (1938) Genetic linkage in graded human characters. Ann Eugenics 8:223–237
Soller M (1992) The evidence for and potential for identification of major single genes effects in trypanotolerant livestock. In: Proc FAO Expert Consultation Genet Aspects Trypanotolerance. FAO, Rome pp 44–76
Soller M, Beckmann JS (1990) Marker-based mapping of quantitative trait loci using replicated progenies. Theor Appl Genet 80: 205–208
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 a cross between inbred lines. Theor Appl Genet 47:35–39
Van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84:803–811
Weller JI (1987) Mapping and analysis of quantitative trait loci in *Lycopersicon* (tomato) with the aid of genetic markers using approximate maximum likelihood methods. Heredity 59: 413–421
Weller JI, Soller M, Brody T (1988) Linkage analysis of quantitative traits in an interspecific cross Tomato (*Lycopersicon esculentum* × *Lycopersicon pimpinellifolium*) by means of genetic markers. Genetics 118:329–339
Weller JI, Kashi Y, Soller M (1990) Power of "daughter" and "granddaughter" designs for mapping of quantitative traits in dairy cattle using genetic markers. J Dairy Sci 73:2525–2532