Genet. Res., Camb. (1999), **73**, *pp.* 61–73. Printed in the United Kingdom

The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species

J. Z. SONG[1], M. SOLLER[1]* AND A. GENIZI[2]
[1] *Department of Genetics, The Silberman Life Sciences Institute, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel*
[2] *Department of Statistics, Agricultural Research Organization, POB 6 Bet Dagan 50250, Israel*

(*Received 7 October 1997 and in revised form 28 May 1998*)

Summary

A full-sib intercross line (FSIL) is constructed in an outcrossing species by mating two parents and intercrossing their progeny to form a large intercross line. For given statistical power, a FSIL design requires only slightly more individuals than an F2 design derived from inbred line cross, but 6- to 10-fold fewer than a half-sib or full-sib design. Due to population-wide linkage disequilibrium, a FSIL is amenable to analysis by selective DNA pooling. In addition, a FSIL is maintained by continued intercrossing so that DNA samples and phenotypic information are accumulated across generations. Continued intercrossing also leads to map expansion and thus to increased mapping accuracy in the later generations. A FSIL can thus provide a bridge to positional cloning of quantitative trait loci (QTL) and marker-assisted selection in outcrossers; and is particularly effective in exploiting the QTL mapping potential of crosses between selection lines or phenotypically differentiated populations that differ in frequency, but are not at fixation, for alternative QTL alleles. In the course of the power analyses, it is shown that for F2 and FSIL designs, power is a function of Nd^2 alone, where N is the total size of the mapping population and d is the standardized gene effect; while for half-sib and full-sib populations, power is a function of Nd^2 and of the number of families included in the mapping population. This provides a convenient means of estimating power for a wide variety of mapping designs.

1. Introduction

Mapping quantitative trait loci (QTL) is a basic operation for positional cloning and for application of marker-assisted selection or marker-assisted introgression in genetic improvement (Soller, 1994). In selfing species, and outcrossers for which inbred lines are available, QTL mapping through linkage analysis with genetic markers is efficiently carried out in F2 or backcross populations (Soller *et al*., 1976). For such species, designs are also available that can increase the efficiency and accuracy of QTL mapping. These include selective genotyping (Lander & Botstein, 1989; Darvasi & Soller, 1992), selective DNA pooling (Darvasi & Soller, 1994; Lipkin *et al*., 1998), development of advanced intercross lines (Darvasi & Soller, 1995), and various procedures that can be classed under the general rubric 'Genetic Chromosome Dissection' (for a review see e.g. Darvasi, 1997). Thus, in these species, one can envision a relatively straightforward route leading from QTL mapping to marker-assisted selection on the one hand or positional cloning on the other. The same holds generally true for dairy cattle, and many species of fish and fruit trees, in which QTL mapping can be carried out within single enormous half-sib or full-sib families.

In the remaining outcrossing species, however, such as poultry, sheep and swine, marker-based linkage mapping of QTL is generally thought of as requiring an accumulation of data over a number of relatively small half-sib or full-sib families (family size in the range 2–100) that together make up the mapping population. In addition, over the mapping population as a whole there is a strong tendency to linkage equilibrium of marker alleles and QTL alleles. Small family size and linkage equilibrium reduce the power of full-sib and half-sib populations as much as 10-fold

* Corresponding author. Telephone: +972 2 6585104. Fax: +972 2 6586975. e-mail: soller@vms.huji.ac.il.

compared with populations derived from crosses between inbred lines (Soller & Genizi, 1978; Weller *et al*., 1990; Knott & Haley, 1992; van der Beek *et al*., 1995; Knott *et al*., 1996). Although mapping accuracy has not been specifically simulated, it is plausible that it too is reduced in proportion. The problem is confounded because the tendency to linkage equilibrium also precludes the utilization of most of the above-mentioned methods for increasing mapping efficiency and accuracy. This poses a major block to linkage mapping and positional cloning of QTL, and to the implementation of marker-assisted selection in these species.

A full-sib intercross line (FSIL) is produced by mating two parents and intercrossing their progeny over two or more generations to form a large 'mapping population'. The genetic make-up of the mapping population is thus completely determined by the chromosome constitution of the two founder parents, and tends, therefore, to a high degree of linkage disequilibrium between marker alleles and QTL alleles. The FSIL design thus provides a QTL mapping population for utilization in outcrossing species, with power and other attributes similar to those obtained for populations derived from a cross between inbred lines. In the present study we analysed the statistical power of a FSIL design, according to the specific marker and QTL composition of the founding parental pair, and compared this with the power of F2, full-sib and half-sib designs. Potential applications of FSIL designs for positional cloning and marker-assisted selection are also discussed.

2. Methods

(i) *The mapping designs*

Inbred line design. An inbred line design has two versions, based on F2 and backcross populations, respectively. In all that follows, only the F2 version will be considered. In this form, a number of parent individuals are taken from each of two fully inbred lines. These are crossed to produce an F1 generation; the F1 individuals are selfed or intercrossed to produce an F2 generation, which serves as the mapping population.

FSIL design. Two parents are chosen, either from the same or different source populations. The parental individuals are mated to produce a large first-generation full-sib family. The full-sibs are intermated at random to produce a first intercross generation. The first intercross generation individuals are again intermated at random to produce the second intercross generation. This procedure is continued indefinitely. Each generation, or any combination of generations from and including the first intercross generation, can serve as a mapping population.

Full-sib design. A series of parental pairs are chosen from a source population. Each parental pair is mated to produce a full-sib family. The entire set of full-sib families serves at the mapping population.

Half-sib design. A series of males are chosen from the same source population. Each is mated at random to a series of females from the same population, to produce a half-sib family. The entire set of half-sib families serves as the mapping population. In the present analysis we consider the situation of a single offspring from each female. This is equivalent to a 'daughter design' (Weller *et al*., 1990).

(ii) *Markers and QTL*

We assume a QTL with two alleles A_1 and A_2, and a marker with many co-dominant alleles, M_1, M_2, M_3, M_4, etc., in complete linkage to the QTL. Genotypic effects at the QTL are $A_1A_1 = d$, $A_1A_2 = h$, $A_2A_2 = -d$, where d and h are in standardized units.

Inbred line design. An inbred line design can uncover marker–QTL linkage only for those marker–QTL pairs for which the parental lines differ in allelic status. Thus, in analysing the inbred line design, the parental line P1 is assumed homozygous for marker–QTL allele combination $M_1M_1A_1A_1$, and the parental line P2 is assumed homozygous for marker–QTL allele combination $M_2M_2A_2A_2$. The F2 mapping population thus consists of three offspring genotypes – $M_1M_1A_1A_1$, $M_1M_2A_1A_2$ and $M_2M_2A_2A_2$ – in proportions 0·25, 0·50 and 0·25, respectively.

FSIL design. The two founder parents of a FSIL mapping population carry a total of four homologues of each chromosome type. For any given set of four parental homologues there are five different archetype marker combinations, in the sense that they have different expectations with regard to number, value and frequency of marker genotypes in the mapping population. These archetype marker combinations are $M_1M_1M_1M_1$, $M_1M_1M_1M_2$, $M_1M_1M_2M_2$, $M_1M_1M_2M_3$ and $M_1M_2M_3M_4$. All other marker combinations are simple permutations or substitutions in one of the above, and have equivalent expectations as their archetypes. For example, $M_1M_1M_1M_1$ and $M_2M_2M_2M_2$ are equivalent; as are $M_1M_1M_1M_2$, $M_1M_1M_2M_1$, and $M_2M_2M_2M_1$, etc. For convenience these five different archetype marker combinations will be denoted M_{1111}, M_{1112}, M_{1122}, M_{1123}, M_{1234}.

In the same sense as for the marker combinations, for a diallelic QTL there are three different archetype QTL combinations among the four parental chromosome homologues. Using the same notation as above, these are denoted A_{1111}, A_{1112} and A_{1122}, respectively. Specific marker–QTL combinations are denoted accordingly; e.g. a set of four parental homologues having composition M_1A_1, M_1A_1, M_1A_2, M_2A_1, respectively, is denoted $M_{1112}A_{1121}$. Note that in the

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

FSIL design, the M_{1122} and A_{1122} combinations include the M_{1212} and A_{1212} variants, but in the full-sib design these must be treated separately (see later).

On the stated assumption of zero recombination between marker and QTL, offspring genotypes in the second and subsequent generations of an FSIL are composed of various combinations of any two of the four parental homologues. On the further assumption of random mating, offspring genotypes in each generation are considered to be distributed with the same Hardy–Weinberg expectations, according to the frequency of the various marker–QTL combinations among the four parental homologues. For example, given the above four parental homologues with composition M_1A_1, M_1A_1, M_1A_2, M_2A_1, their respective frequencies are: $M_1A_1 = 0{\cdot}50$ and $M_1A_2 = M_2A_1 = 0{\cdot}25$.

The FSIL design can uncover marker–QTL linkage only for those parental pairs that are in linkage disequilibrium for markers and QTL. This excludes monomorphic marker and QTL combinations, and also certain specific marker–QTL combinations such as $M_{1122}A_{1212}$, or variants thereof, that are in linkage equilibrium. Thus, some of the QTL segregating in a source population will not be uncovered by a FSIL design based on a pair of individuals chosen from that population. Given a number of markers in each chromosomal region, however, at least one or more can be expected to be both polymorphic and in linkage disequilibrium with a QTL in that region that is polymorphic in the parental pair. Thus, the above limitation applies specifically to QTL that are monomorphic in the parental pair.

Full-sib design. The two founder parents of each family of a full-sib design mapping population also carry a total of four homologues for each chromosome type. Offspring genotypes within each family are assumed to be distributed according to elementary genetic expectations. In this design, as noted above, there are six different archetype marker combinations and four different archetype QTL combinations, since the M_{1122} and A_{1122} combinations are split into M_{1122} and M_{1212}, and A_{1122} and A_{1212} combinations, respectively, which differ in their expectations with respect to number, value and frequency of genotypes in the mapping population. As in the FSIL design, only those parental pairs that are in linkage disequilibrium for markers and QTL are informative for marker–QTL linkage. This excludes the same marker or QTL chromosome combinations as in the FSIL design. In addition, the first-generation offspring must be segregating, which also excludes the M_{1122} or A_{1122} parental combinations, or variants thereof.

In contrast to the FSIL design, however, while any individual family may have a marker–QTL composition that is uninformative with respect to marker–QTL linkage, the collection of full-sib families provides information on all segregating QTL in the source population (unless one of the alleles at the QTL is close to fixation, or the number of full-sib families in the experiment is small). Here too, in principle, proper choice of markers could also bring all informative QTL combinations into play in the full-sib family design, in this way increasing power. However, in practice and theory this would be much more difficult, since markers would need to be adjusted separately for each family.

Half-sib design. Only instances in which the sire of the half-sib family is heterozygous at both marker and QTL are informative for marker–QTL linkage. In the mapping population offspring genotypes are assumed to be distributed according to expectation assuming Hardy–Weinberg equilibrium at the QTL in the dam population. Here too, although any sire homozygous at the QTL is uninformative with respect to marker–QTL linkage, the experiment as a whole will be informative for all QTL segregating in the source population, subject to the same exceptions noted for the full-sib family design.

(iii) *The marker combinations*

Inbred line and half-sib designs. In both these designs the only requirement is marker heterozygosity. Since a wealth of highly polymorphic markers is available for QTL mapping purposes, we assume the desired condition for all markers.

FSIL and full-sib designs. In both these designs, power of the mapping population will depend on the specific marker combinations present. The probability of obtaining any particular marker combination is given by a multinominal distribution, and depends on the marker composition of the population from which the parents are drawn. For the general case, where marker allele frequencies are unequal, the expected sampling probability, P_m, of the various archetype marker combinations is given as sums of multinominal probabilities corresponding to the individual marker combinations, $C(m)$ that belong to archetype combination m. P_m can be written in the form

$$P_m = \sum_{C(m)} [4!/\Pi x_i!]\,\Pi p_i^{xi},$$

where p_i is the frequency of the ith marker allele M_i in the source population; x_i is the number of parental chromosomes carrying marker allele, M_i, $\Sigma x_i = 4$; and $\sum_{C(m)}$ indicates summation over all combinations that constitute the archetype combination, m.

When all $p_i = p$ are equal, so that $p = 1/R$, the expression for P_m simplifies to

$$P_m = kC_r^R[4!/\Pi x_i!](1/R^4),$$

where R is the total number of different marker alleles in the source population; r is the number of different

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

marker alleles among the four parental chromosomes in marker combination, m; and k is the number of permutations relevant to this combination. Thus, $k = 1$ for marker combinations M_{1111}, M_{1122} and M_{1234}; $k = 2$ for marker combination M_{1112}; and $k = 3$ for marker combination M_{1123}.

(iv) *The QTL combinations*

Inbred line design. When the two inbred parental lines are derived from the same source population, the probability that the two lines differ at the QTL is equal to $2pq$. Ordinarily, however, in the case of an inbred line design the two parental inbred lines will have been derived from different source populations. In this case the proportion of heterozygous QTL combinations will depend on the genetic distance between the source populations. Since this will vary and is generally unknown, all calculations are on the assumption that the parental lines differ in QTL allele status.

FSIL design. When the two parental individuals are taken from the same source population, the probability of any particular QTL combination is given by $p^{x_1}q^{x_2}$, where p and q are frequencies of QTL alleles A_1 and A_2 in the source population; and x_1 and x_2 $(= 4-x_1)$ are the number of parental chromosomes carrying QTL alleles A_1 and A_2, respectively. When the two parental individuals are each taken from a different source population, the probability of any particular QTL combination is given by $p_1^{x_{11}}q_1^{x_{12}}p_2^{x_{21}}q_2^{x_{22}}$, where p_1 and p_2 denote the frequency of allele A_1 in source populations 1 and 2, respectively; x_{11} and x_{12} $(= 2-x_{11})$ refer to the number of chromosomes of the parent taken from source population 1, that carry alleles A_1 or A_2 respectively; and x_{21} and x_{22} $(= 2-x_{21})$ are the corresponding values for the parent taken from source population 2. Calculations of QTL combination frequencies were carried out for three situations: both parents from the same source population, with equal allele frequencies at the QTL $(p = q = 0{\cdot}5)$; both parents from the same source population, with unequal allele frequencies at the QTL $(p = 0{\cdot}8,\ q = 0{\cdot}2)$; and each parent from a different source population, with different allele frequencies (source population 1: $p = 0{\cdot}8$, $q = 0{\cdot}2$; source population 2: $p = 0{\cdot}2$, $q = 0{\cdot}8$).

Full-sib design. In this design, the two parental individuals are always taken from the same source population. The probability of any particular parental QTL combination is thus given by the corresponding expressions of the FSIL design. Since a full-sib family design will generally include a large number, k, of families $(k > 20)$, it is assumed that the distribution of the various QTL combinations within any given marker combination is in proportion to their probabilities. Calculations were for $p = q = 0{\cdot}5$, which maximizes the number of informative families. Clearly any other situation will increase the proportion of uninformative parent combinations and decrease power. Similarly, the proportion of uninformative parent combinations is increased by taking each parent from a different source population, except for the trivial case when the two source populations have identical allele frequencies. Producing the parents by crossing two separated populations, however, will often increase the proportion of heterozygous parents and increase power.

Half-sib design. The probability that any given sire is heterozygous at the QTL is equal to $2pq$. When the total number of families is large $(k \geqslant 20)$, it can be assumed that this is indeed close to the proportion of heterozygous sires in any given experiment. When the total number of families in the experiment is small, however $(k < 20)$, the proportion of heterozygous sires in any given experiment will vary widely, according to a binomial distribution with parameters k and $2pq$. Calculations were on the best case basis, $p = q = 0{\cdot}5$, which maximizes the number of informative sires. Here, too, deviation of allele frequencies from equality will increase the proportion of uninformative homozygous sires and decrease power; while producing the sires by crossing two separated populations will increase the proportion of heterozygous sires and increase power.

(v) *Power calculations*

Henceforth, to simplify notation, M_i and A_h will be used for marker and QTL combinations, rather than for the state at a single allele; accordingly, M_iA_h will denote a marker–QTL combination. To test for linkage in a given mapping population derived from an M_iA_h marker–QTL combination, one-way ANOVA with significance level α is used among the different marker genotype groups. Power of the test is defined as equal to $1-\beta$, where β is the Type II error of the experiment. The Type II error is given by a non-central F distribution (Scheffe, 1959), as follows:

$$\beta = \text{Prob}\, F(n_1, n_2, \Psi) < F(n_1, n_2, \alpha),$$

where n_1, d.f. in numerator $= m_i - 1$; n_2, d.f. in denominator $= N - m_i$; and Ψ, the non-centrality parameter $= N\Sigma w_{ij}(\mu_{ijh} - \mu_h)^2$. In turn m_i is the number of different marker genotypes in a mapping population or family originating from the M_i parental marker combination; N is the total number of individuals in the mapping population or family derived from the M_iA_h marker/QTL combination; w_{ij} is the frequency of marker genotype G_j within the mapping population originating from parental marker combination M_i; μ_{ijh} is the genotypic value of marker genotype G_j,

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

Table 1. *Parameters for noncentral F power calculations according to mapping design: n_1, d.f. in the numerator; n_2, d.f. in the denominator; Ψ, noncentral parameter*

Design	n_1	n_2	Ψ
Inbred line	2	$N-3$	$N\Sigma w_j(\mu_j-\mu)^2$
FSIL[a]	m_i-1	$N-m_i$	$N\Sigma w_{ij}(\mu_{ijh}-\mu_h)^2$
Full-sib[b]	$k(m_i-1)$	$N-km_i$	$N\Sigma v_h\Sigma w_{ij}(\mu_{ijh}-\mu_h)^2$
Half-sib[c]			
$k>10$	k	$N-k$	$2pqN\Sigma(\mu_j-\mu)^2/2$
$k\leqslant 10$	k	$N-k$	$H_xN\Sigma(\mu_j-\mu)^2/2$

[a] Calculated for given marker composition, M_i; separately for each QTL composition, A_h within M_i. Power was calculated as $\Sigma v_h(1-\beta_h)$, where $1-\beta_h$ is power for A_h within M_i and v_h is frequency of A_h *within* M_i, including only A_h that are informative for QTL mapping.

[b] Calculated for given marker composition, M_i; over all possible QTL compositions, A_h within M_i.

[c] Calculated over homozygous and heterozygous QTL compositions. For this design there are always two marker genotype classes in the mapping population, each at frequency 0·5. For $k>10$, $2pq$ was taken as the expected proportion of sires, heterozygous at the QTL. For $k\leqslant 10$, the non-central parameter and power, β_x, were calculated separately for each possible proportion of sires, H_x out of k, heterozygous at the QTL. Expected power was then calculated as $\Sigma v_{H_x}(1-\beta_x)$, where v_{H_x} is the binomial probability of finding a proportion H_x of sires heterozygous at the QTL in a sample of k sires.

Notation: N = total number of individuals in the mapping population; m_i = number of different marker genotypes in a FSIL design mapping population, or full-sib family, originating from the M_i parental marker combination; w_j = frequency of marker genotype G_j in an F2 population; μ_j = genotypic value of marker genotype G_j in an F2 population; μ = mean of F2 population; w_{ij} = frequency of marker genotype G_j within FSIL mapping population or full-sib family originating from parental marker combination M_i; μ_{ijh} = genotypic value of marker genotype G_j, within parental marker combination M_i, and parental QTL combination A_h; μ_h = mean genotypic value of FSIL mapping population or full-sib family originating from A_h; k = the number of full-sib or half-sib families in the mapping population; v_h = the proportion of full-sib families having QTL combination A_h; p and q, frequency of QTL alleles A_1 and A_2, respectively, in the source population.

within parental marker–QTL combination M_iA_h; and μ_h is the mean genotypic value of mapping population originating from parental QTL combination A_h.

Specific details showing how this expression was adapted for the various mapping designs are given in Table 1.

The marker combinations. In principle, given multi-allelic markers and saturated genome maps, the parental chromosomes can be genotyped for numerous markers in each chromosomal region, until a desired marker combination is found. For the inbred line and half-sib family designs there is only a single informative marker combination. Namely, for the inbred line design the two parental lines must be homozygous for alternative alleles at the marker being analysed; and for the half-sib family design the sire of the half-sib family must be heterozygous for the marker being analysed. In both cases it is assumed that the required condition is met for all markers and (in the case of the half-sib family design) all sires being analysed. For the chromosome and full-sib family designs, however, there are a number of different informative marker combinations. In these cases, each marker combination was treated separately in the power analyses. This was done to provide some indication of the benefit that would be obtained by pursuing a more informative marker combination in the relevant chromosomal region.

The QTL combinations. In contrast to the situation with respect to markers, the specific QTL chromosomal combination in the parental chromosomes of any particular mapping population (for inbred line or FSIL design) or family (for full-sib or half-sib family designs) is in principle unknown; yet the various combinations differ in power and some have zero power. This has different consequences, depending on whether an experiment derives from a single parental pair (as in inbred line or FSIL designs) or from a number of parents or parental pairs (as in half-sib and full-sib family designs).

In the inbred line design, a specific experiment has only two possible outcomes with respect to a specific QTL: maximum power with respect to all QTL for which the parental lines differ in allelic state, and zero power with respect to all QTL for which the parental lines are similar in allelic state. Thus, any given experiment will be informative only with respect to some fraction of the QTL segregating in the original source populations from which the inbred lines were derived. However, with respect to the QTL for which the cross is informative, the fact that some other cross might be uninformative with respect to these QTL does not decrease power in any way. Thus, for the inbred line design, power calculations are based on the assumption that the parental lines differ at the QTL.

In the FSIL design, depending on the parental QTL combination, a particular experiment can have a variety of outcomes with respect to a specific QTL. These differ sharply in power, and include zero power for the case where the two parents are monomorphic or in linkage equilibrium with respect to the QTL. Here too, if a given FSIL design mapping population is informative, the fact that a different FSIL design mapping population might be less informative, or even uninformative at this marker, does not reduce power in any way. Thus, an average power value would not represent the actual power for any experiment. Consequently, for the FSIL design, we provide separate power values for each of the informative QTL combinations.

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

In the full-sib family design, a mapping population will generally consist of a fairly large number of families ($k > 20$). In this case, any particular experiment will not vary widely from the expected proportions of the various types of parental QTL combinations. Thus, a satisfactory approach is simply to calculate the expected noncentrality parameter over the experiment as a whole, according to the probability of each parental QTL combination, and then obtain the overall power of the experiment for this non-centrality parameter.

In the half-sib design, the same holds true when a large number of half-sib families are included in the mapping population, in which case the proportion of sires heterozygous at the QTL will not vary widely. When the mapping population comprises only a small number of half-sib families, however, the proportion of sires heterozygous at the QTL can vary widely from one experiment to another, according to a binomial sampling distribution. In this case, we followed the procedure of Weller *et al*. (1990), and calculated the expected power of each configuration separately, and then calculated the mean expected power, weighted according to the probability of each configuration of sire QTL genotypes.

Note that in the case of full-sib and half-sib families, the uninformative families are an integral part of each and every experiment. Thus, the uninformative families necessarily reduce the power of each experiment and must be included in power calculations. The positive aspect of this situation is that once the number of families in an experiment is greater than 10, there is essentially zero probability, given that the alternative QTL alleles are at moderate frequencies, of all families in a given experiment being uninformative with respect to a given QTL. Thus, in contrast to inbred line or FSIL designs, a single full-sib or half-sib experiment can potentially access all QTL segregating in the source population.

Table 1 gives n_1, n_2 and Ψ as calculated for the various designs, as a function of total population size (N) and, for the full-sib and half-sib designs, the number of families (k) within the mapping population. Examination of the expressions for the non-centrality parameter, Ψ, show that, on the assumption that $h = 0$, they are all functions of N and of some variant of $(\mu_i - \mu)$; where μ_i is the genetic value of a marker genotype and μ is the population or family mean. Both μ_i and μ are expressed in units of QTL gene effect, d, and dominance h. Hence, assuming $h = 0$, it is evident that the non-centrality parameters can all be expressed in the form $Nd^2\delta^2$, where δ^2, denoted the 'design parameter', is a function of the experimental design and of marker and QTL composition. This means that within a particular design, all combinations of population size and gene effect that yield the same value of Nd^2 will have the same power. Furthermore, since δ^2 is a known function of the design, depending on gene frequencies, the non-centrality parameter of a design can readily be calculated for any combination of values of N and d.

In order to provide a convenient basis of comparison of the various designs, Nd^2 values for power from 0·10 to 0·90 are provided for the inbred line design. Relative Nd^2 values, denoted RN values, are given for all other designs as a multiple of the inbred line Nd^2 values necessary to obtain equivalent power.

3. Results

Table 2 shows the probability of the various marker combinations for the two parents of the FSIL or full-sib family designs, according to number, R, of marker alleles in the population from which the parent individuals are taken, assuming equal frequency for all marker alleles, $m_i = 1/R$. When $R = 2$, a proportion of 0·875 of markers will be found in informative combinations for the FSIL design, and 0·750 will be informative for the full-sib family design. When $R = 3$, these proportions improve dramatically, to 0·963 and 0·889 for the FSIL design and full-sib design, respectively. When $R \geqslant 4$, over half the marker

Table 2. *The probability of the various marker combinations for the two parents of the FSIL or full-sib family designs, according to number, R, of marker alleles in the population from which the parent individuals are taken, assuming equal frequency for all marker alleles, $p_i = 1/R$*

Marker combination	R					
	2	3	4	5	8	10
1111	0·125	0·037	0·016	0·008	0·002	0·001
1122[a]	0·125	0·074	0·047	0·032	0·014	0·009
1212	0·250	0·148	0·094	0·064	0·027	0·018
1112	0·500	0·296	0·188	0·128	0·055	0·036
1123[b]	—	0·148	0·188	0·192	0·164	0·144
1213	—	0·296	0·375	0·384	0·328	0·288
1234	—	—	0·094	0·192	0·410	0·504

[a] For FSIL design 1122 and 1212 combinations are equivalent, and are referred to in the text as the M_{1122} marker combination. Probability of the archetype 1122 marker combination was calculated using the expression for P_m. This was then split into 1122 and 1212 combinations by noting that within the 1122 archetype, combinations of the form 1212 are twice as frequent as combinations of the form 1122.

[b] For FSIL design 1123 and 1213 combinations are equivalent, and are referred to in the text as the M_{1123} marker combination. Probability of the archetype 1123 marker combination was calculated using the expression for P_m. This was then split into 1123 and 1213 combinations by noting that within the 1123 archetype, combinations of the form 1213 are twice as frequent as combinations of the form 1123.

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

Table 3. *FSIL design. The number of permutations* (P), *the design parameter* (δ^2), *and the frequency of each QTL combination within each marker combination, according to source population*[a]

Marker combination	QTL combination	P	δ^2	Source population		
				Equal	Unequal	Cross
1112	1111	2	0	0·125	0·411	0·051
	1112	2	0·375	0·125	0·109	0·109
	1121	6	0·043	0·375	0·326	0·326
	1122	6	0·165	0·375	0·154	0·514
1122	1111	2	0	0·125	0·411	0·051
	1112	8	0·125	0·500	0·435	0·435
	1122	2	0·500	0·125	0·051	0·411
	1212	4	0	0·250	0·103	0·103
1123	1111	2	0	0·125	0·411	0·051
	1112	4	0·375	0·250	0·218	0·218
	1211	4	0·125	0·250	0·218	0·218
	1122	2	0·500	0·125	0·051	0·411
	1212	4	0·250	0·250	0·102	0·102
1234	1111	2	0	0·125	0·411	0·051
	1112	8	0·375	0·500	0·435	0·435
	1122	6	0·500	0·375	0·154	0·514

[a] See text for details.

combinations will be of the highly informative M_{1123} or M_{1234} types.

Table 3 shows the number of combinations and permutations of each QTL combination within marker combinations for the FSIL design. Also shown are the design parameter for each QTL combination and the frequencies of each QTL combination within each marker combination according to source population. The design parameter varies quite widely according to QTL combination within marker combinations, but there is a general tendency for higher design parameters in the M_{1123} and M_{1234} marker combinations than in the M_{1112} and M_{1122} marker combinations. The frequency of the uninformative A_{1111} QTL combination is much greater when the QTL allele frequencies of the source population are unequal; while the frequency of the highly informative A_{1122} QTL combination is much increased when the two parents of the FSIL design derive from source populations that differ widely in QTL allele frequencies.

Table 4 shows the same parameters for the full-sib family design. The design parameter again varies quite widely according to QTL combination within marker combinations. In addition, the proportion of uninformative QTL combinations also varies within marker combinations, decreasing from the M_{1112} and M_{1123} combinations, through the M_{1212} to the M_{1213} combinations. Again, the frequency of the uninformative A_{1111} QTL combination is much greater when the QTL allele frequencies of the source population differ widely. In marked contrast to the FSIL design, for the full-sib design the frequency of the uninformative A_{1122} QTL combination also increases when the two parents are taken from different source populations (not shown). If the two parents of the full-sib families are produced by crossing source populations that differ widely in QTL allele frequency, however, the frequency of the informative A_{1212} QTL combination increases.

Overall design parameters are as follows. *For the inbred line cross*, $\delta^2 = 0{\cdot}50$. *For the half-sib design*, the overall half-sib design parameter is $pq/2$, hence, $\delta^2 = 0{\cdot}250$ when the sire is heterozygous at the QTL, and 0 otherwise. *For the full-sib and FSIL designs*, overall design parameters for the FSIL design were calculated by weighting the design parameters of the various QTL combinations by their frequency, assuming $p = q = 0{\cdot}5$. *For the full sib design*, $\delta^2 = 0{\cdot}125, 0{\cdot}125, 0{\cdot}125$ and $0{\cdot}250$ for marker combinations M_{1112}, M_{1123}, M_{1212} and M_{1234}, respectively. *For the FSIL design*, for comparison with the inbred line design, it is instructive to include informative QTL combinations only. In this case, we obtain $\delta^2 = 0{\cdot}142, 0{\cdot}200, 0{\cdot}286$ and $0{\cdot}429$ for marker combinations M_{1112}, M_{1122}, M_{1123} and M_{1234}, respectively. For comparison with the full-sib design it is instructive to include the uninformative A_{1111} QTL combination as well, reducing the above values to $0{\cdot}125$, $0{\cdot}125$, $0{\cdot}250$ and $0{\cdot}375$, respectively. The overall superiority of the inbred line cross to all other designs is expressed clearly in its greater design parameter, equal to $0{\cdot}500$. This value is approached, however, by the $M_{1234}A_{1122}$ combination in the FSIL design, and is equalled by specific marker–QTL

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

Table 4. *Full-sib design. The number of permutations* (P), *the design parameter* (δ^2), *and the frequency of each QTL combination within each marker combination, according to source population*[a]

Marker combination	QTL combination	P	δ^2	Source population	
				Equal	Unequal
1112	1111	2	0	0·125	0·411
1123	1112	4	0·250	0·250	0·218
	1211	4	0	0·250	0·217
	1212	4	0·250	0·250	0·102
	1122	2	0	0·125	0·051
1212	1111	2	0	0·125	0·412
	1112	8	0·125	0·500	0·435
	1212	2	0·500	0·125	0·051
	1221	2	0	0·125	0·051
	1122	2	0	0·125	0·051
1234	1111	2	0	0·125	0·412
1213	1112	8	0·250	0·500	0·435
	1212	4	0·500	0·250	0·102
	1122	2	0	0·125	0·051

[a] See text for details.

Table 5. *Inbred line design. Power as a function of* Nd^2

Nd^2	Power
4·63	0·10
10·00	0·27
16·48	0·50
20·00	0·61
35·20	0·90
40·00	0·94
90·00	1·00

N, population size; d, gene effect at the QTL.

combinations within the above archetype. The advantage of the FSIL design relative to the full-sib design, even when uninformative QTL combinations are included in the FSIL design, derives primarily from the A_{1122} QTL combination, which is uninformative in the full-sib design but highly informative in the FSIL design. Note that, in the case of the FSIL and half-sib designs, the overall design parameters were not used for the actual power calculations. As noted, the power calculations were based on the weighted mean of the power of the various QTL combinations, and not on the weighted mean of their non-centrality parameter.

Table 5 shows power as a function of Nd^2 for the inbred line design, at Type I error, $\alpha = 0{\cdot}01$. Powers of 0·10, 0·50 and 0·90 require Nd^2 values of 4·63, 16·48 and 35·2, respectively. As an example of the use of Nd^2 values, we note that from the given Nd^2 value for power of 0·90, it follows that this power will be obtained at $N = 100$, $d = 0{\cdot}59$; $N = 500$, $d = 0{\cdot}27$; and $N = 1000$, $d = 0{\cdot}19$.

Table 6. *FSIL design. Relative* Nd^2 (*RN*) *required with the FSIL design to provide the same power as the inbred line design, according to marker and QTL composition*

Marker composition	QTL composition	Power		
		0·10	0·50	0·90
1112	1112	1·46	1·34	1·33
	1121	12·48	11·75	11·64
	1122	3·26	3·06	3·02
1122	1112	4·09	4·00	3·98
	1122	1·00	1·00	1·00
1123	1112	2·02	1·79	1·68
	1211	6·09	5·38	5·07
	1122	1·64	1·35	1·26
	1212	3·03	2·69	2·53
1234	1112	2·66	2·22	2·00
	1122	2·02	1·66	1·50

Table 6 shows the relative Nd^2 (RN) required with the FSIL design to provide the same power as the inbred line design, according to marker and QTL composition. On the assumption of zero recombination between marker and QTL, these RN values will hold for all generations of the FSIL design. As an example of the use of the RN values, we note that for the $M_{1112}A_{1112}$ marker/QTL combination, the RN value for a power of 0·90 is 1·33. This means that the Nd^2 value required for a power of 0·90 with this combination is 1·33-fold that required for equivalent power by an inbred line design, i.e. $1{\cdot}33 \times 35{\cdot}20 =$

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

46·8; from which it follows that for a QTL of effect 0·27, a population size of $N = 46{\cdot}8/(0{\cdot}27)^2 = 642$ is required for power of 0·90, as compared with $N = 500$ for the inbred line design. RN values tend to be a little smaller for power of 0·9 than for power of 0·1, but are within $\pm 10\%$ of the values for power 0·50 across the power range. Marker composition plays a major role in determining RN, creating up to a 9-fold difference in RN for QTL composition A_{1112} (i.e. any parental QTL set with three alleles of one type and the fourth of a different type); and up to a 3-fold difference in RN for QTL composition A_{1122}. The M_{1234} marker composition, however, delivers consistently low RN values. Taken over all QTL compositions, most RN values are in the range 1·0 to 4·0 at any given power, with a modal value of 2·2. Thus, it seems reasonable to conclude that, on average, a FSIL design will deliver power equivalent to that of an inbred line design, at twice the corresponding population size or Nd^2.

However, this 'average' view does not exhaust the possibilities of the FSIL design. The reason for this is that some of the marker combinations can be transformed into other more powerful marker combinations, according to the specific marker–QTL coupling relationships in the parental chromosomes. For example, marker/QTL combination $M_{1112}A_{1121}$ has an RN of 11·64 for power of 0·90; while marker/QTL combination $M_{1112}A_{1112}$ has an RN of 1·33. Yet by judicious search, a marker having distribution M_{1121} A_{1121} can be found, which will provide an RN of 1·33. Similarly, marker combination $M_{1234}A_{1112}$ has an RN of 2·00. Yet, by combining marker alleles having similar effects in the analysis, it should be possible to transform this into an $M_{1112}A_{1112}$ combination with an RN of 1·33. In general, by judicious choice of markers or by combining marker alleles having similar associated effects, it should be possible to transform any parental pair having an A_{1112} QTL composition into an $M_{1112}A_{1112}$ combination, with an RN of 1·33; and any parental pair having an A_{1122} QTL composition into an $M_{1122}A_{1122}$ combination with an RN of 1·00. Thus, in the analysis of an FSIL design, one would initially use a fairly high Type I error, to identify potential marker–QTL linkages, and then subject these to a second round of marker search and analysis, attempting to identify a marker combination that greatly increased statistical significance. In this way, it should be possible to obtain from a FSIL design population power virtually equivalent to that of an inbred line design.

Table 7 shows the RN required with the half-sib design to provide the same power as the inbred line design, according to number of sires tested. RN values increase strongly with the number of sires tested. This reflects the change in sire configurations in moving from 5 to 20 sires, and also the fact that with more sires tested, fewer daughters are tested per sire. RN values for this design do not vary widely, remaining in the range of 4 to 10, with a modal value of about 6.

Table 7. *Half-sib design. Relative Nd^2 (RN) required with the half-sib design to provide the same power as the inbred line design, according to number of sires tested within a fixed total population size, N, assuming equal frequency for the two QTL alleles*

No. of sires	Power		
	0·10	0·50	0·90
5	5·88	5·42	4·89
10	7·74	6·70	6·00
20	10·94	8·62	7·53

Table 8. *Full-sib design. Relative Nd^2 (RN) required with the full-sib design to provide the same power as the inbred line design, according to marker composition, and number of families tested within a fixed total population size, N, assuming equal frequency for the two QTL alleles*

Marker composition	No. of families	Power		
		0·10	0·50	0·90
1112	25	9·7	7·9	7·0
1123	50	14·1	10·8	8·8
	100	19·5	14·4	11·9
	200	27·5	19·4	15·6
1212	25	11·4	8·9	7·8
	50	16·9	12·6	9·9
	100	23·9	16·6	13·6
	200	32·9	22·6	17·8
1213	25	6·2	4·8	4·1
1234	50	9·1	6·1	5·6
	100	12·5	8·9	7·2
	200	18·2	12·1	9·6

Table 8 shows the relative Nd^2 (RN) required with the full-sib design to provide the same power as the inbred line design. Extensive calculations (not given) showed that, for a fixed number of families, Nd^2 values for given power were hardly affected by number of full-sibs per family, so long as this number was not less than 5. For example, with 25 families and marker composition M_{1112} or M_{1123}, Nd^2 required for a power of 0·50 with family size of 10, 20 or 40 was 130·6, 126·0 and 122·8, respectively. Similarly, with 50 families, Nd^2 required for a power of 0·50 with family size of 5, 10 or 20 was 178·3, 171·4 and 164·2, respectively. Thus, in both cases, power was affected in the same manner by equivalent proportionate changes in family size (and hence in N) or in d^2. When family size drops below 5, however, degrees of freedom for the denominator drop precipitously, reducing power.

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

The relative independence of Nd^2 for given power on family size, as long as it is 5 or more, enables RN values for the full-sib family case to be presented in Table 8 as a function of number of families only, within marker composition classes. Values of Nd^2 for given number of families in the table are those for the family size closest to, but not less than, 5. Within a given number of families, RN values decrease strongly as a function of power. Although a slight to moderate decrease is observed with the other designs, the effect in the full-sib design is much greater. A reason for this behaviour is not immediately evident. RN values for given power increase strongly with the number of families tested. Thus, RN values for 25 families for a power of 0·50 range from 4·8 to 7·9, depending on marker composition; while those for 200 families at the same power range from 12·1 to 19·4. This reflects the fact that as more families are tested at given total population size, the individual family is necessarily smaller.

In the full-sib design, marker compositions M_{1112}, M_{1123} and M_{1212} have very similar RNs; those for marker compositions M_{1213} and M_{1234} are distinctly smaller. Thus, here too, it seems reasonable to conclude that by judicious choice of markers for each family, a full-sib design could potentially be brought to an RN of 6 to 8. However, this would clearly be a much more onerous task than for a FSIL design, since a separate marker search would need to be carried out for each family. In addition, with small family sizes, chance alone would yield high test statistics with some marker arrangements, not necessarily reflecting the underlying marker–QTL composition. This would inflate test statistics, producing a higher Type I error. Thus effective RN for a full-sib family design may be closer to 10 overall, depending however on marker composition and number of families tested.

All in all, it seems reasonable to conclude that a FSIL design, following adjustment of markers to the optimum, will deliver power equivalent to that of a half-sib or full-sib design at one-sixth and one-tenth the corresponding population size, respectively.

4. Discussion

The results of the present analysis show that in principle the FSIL design brings to QTL mapping in outbreeding populations (animals, fruit trees, some plants) many of the advantages that the inbred line designs provide for selfing species, or species for which inbred lines are available. Foremost among these advantages is the greater statistical power for given population size and gene effect, relative to the alternative full-sib or half-sib designs applicable in outbreeding populations. A useful attribute of the FSIL design is the ability to achieve a major increase in power, in the same sample and data base, by searching for more informative marker combinations in chromosomal regions showing indications of QTL presence. As noted in Section 3, this is not relevant for the half-sib design, and impractical and theoretically complex for the full-sib design. Following marker adjustment, a FSIL design can consistently provide 6- to 10-fold greater power than equivalent half-sib of full-sib family designs. In addition, FSIL mapping populations, being in massive linkage disequilibrium, are amenable to mapping through selective DNA pooling (Darvasi & Soller, 1994; Lipkin *et al*., 1998). This can provide major reductions in genotyping costs of the initial genome scan for QTL of interest.

Yet another advantage of the FSIL design is the ability to continue the population for a number of generations allowing phenotyping and genotyping data to be accumulated over the generations – a matter which can be of practical importance in experimental situations where rearing facilities are limited. A potential drawback of the FSIL design is the fact that the first intercross population is two generations removed from the founder parents. This increases the proportion of recombination between markers and QTL, as compared with the other mapping designs that are only a single generation removed from their founders. Continuing a FSIL population over a number of generations by repeated intercrossing will further increase the cumulative number of recombination events between closely linked sites. Consequently, a FSIL design requires a more dense marker spacing for equivalent power compared with the other designs. This presupposes working with a high density of markers, but is easily accommodated by the use of selective DNA pooling for the initial QTL scan, as described above.

All the outcrosser designs have maximum power with respect to QTL present at moderate allele frequencies: full-sib and half-sib designs access all QTL present at moderate frequencies, while the FSIL design can be expected to access 80–90% of these QTL, depending on their specific allele frequencies. The designs differ more in their power with respect to rare alleles. A full-sib design based on a moderate number of offspring per family will have very low power for QTL present at low frequencies. A half-sib design based on a small number of very large families has greater likelihood of identifying rare alleles, because of the high power of a single large half-sib family (Lipkin *et al*., 1998). The FSIL design is intermediate. It will have high power for a rare allele included by chance among the four parental chromosomes, but only a small proportion of rare alleles will be included Thus, a favourable QTL present at a frequency of 0·01 in the source population would never be identified by a full-sib mapping population based on small to moderately small families; would be

https://doi.org/10.1017/S0016672398003486 Published online by Cambridge University Press

identified in up to 10% of half-sib mapping populations based on, say, 10 very large half-sib families; and would be identified in 4% of FSIL populations. The corresponding values for a favourable QTL allele present at a frequency of 0·1, are: almost never, almost always, and one-third.

In principle, the FSIL design is perfectly general, and does not require prior development or availability of special lines. Thus it can be applied to any mapping situation involving outcrossing populations. The FSIL design provides an especially powerful means of tapping the mapping potential of selection lines, or differentiated populations, which differ in allele frequency at the relevant QTL but are not yet at fixation for alternative alleles. In this case, F2 analyses which assume fixation for alternative alleles (Beckmann & Soller, 1988; Haley *et al*., 1994) lose much of their power. Within-population mapping, by full-sib or half-sib design, will also be relatively ineffective due to the likely shift of QTL allele frequencies to the 'unequal' situation. Half-sib or full-sib design mapping by F1 parents produced by crossing the two lines, however, will be more effective than within-population mapping, because of the high expected proportion of heterozygotes at the QTL.

In practice, the FSIL design has a number of limitations that restrict its field of application. Foremost among these is the fact that it requires developing a dedicated mapping population, outside of the routine commercial breeding procedures. This is probably an insuperable obstacle for its utilization for any of the larger livestock species with low reproductive potential (cattle, sheep and goats); but it should not preclude its utilization for species with high reproductive potential and short generation time, such as mice, rats, fish, poultry, swine, and outcrossing plant species including some trees.

A major practical problem in implementing a FSIL design is the inbreeding associated with production of the F2 generation of the mapping population, since this is produced by intermating the F1 full-sibs born of the initial parental pair. The deleterious effects of the resultant inbreeding might best be circumvented by initiating a number of independent F1 families from independent parental pairs drawn from the source population; intermating each of these F1 families; and continuing with the F2 or F3 families that show least signs of inbreeding depression. Aside from the inescapable inbreeding in the F1 generation, inbreeding in later generations is much reduced simply by maintaining a large effective number of parents, and can be further minimized by deliberate choice of parental mating pairs to avoid inbreeding. Nevertheless, when implementing a FSIL design, it should be kept in mind that some effects mapped in the early generations may be due to rare deleterious recessive alleles that have been moved to relatively high frequency by their chance presence in the founder parental pair. This can be examined by noting changes in associated marker allele frequencies across generations. Some QTL effects may also be modified by interaction with the specific genetic background of the FSIL, or by interaction with more general inbreeding effects.

Continuing an FSIL design population over a number of generations by recurrent intercrossing results in an accumulation of recombination events and map expansion. Thus, for the purposes of fine mapping, an advanced generation FSIL population can provide some of the same advantages as an advanced intercross line (AIL; Darvasi & Soller, 1995). When data and DNA samples are retained from early generations of the cross, these can be used for gross QTL mapping; while data and DNA samples from the later generations are used for fine mapping. The accumulation of recombination events within an advanced-generation FSIL population also lends itself naturally to ultra-fine mapping methods based on analysis of recombinant chromosomes (referenced in Darvasi, 1997). In this way, a FSIL population can provide a basis for positional cloning of QTL.

A FSIL design can also be viewed as a procedure for general identification and mapping of QTL, not necessarily related to a specific commercial breeding population. In this case, the information on QTL effects and map location obtained from the FSIL design can feed into an MAS programme in a commercial population in a number of ways: (i) By focusing QTL mapping efforts in the commercial population on chromosomal regions that are likely to contain QTL. (ii) By serving as a step to linkage disequilibrium QTL mapping in the commercial population. Specifically, because of the mapping precision possible with a FSIL design, it is possible to saturate the region containing the QTL with markers, aiming at a spacing of 1–2 cM (so that on the average a QTL would be only 0·25–0·50 cM from the nearest marker) and then look for marker-associated effects on a population-wide basis (Lande & Thompson, 1990). (iii) By serving as a source of 'marked' QTL for introduction into the commercial line by marker-assisted introgression. In particular, as in an inbred line design, a FSIL design provides information on specific marker–QTL phase relationships. Consequently, once a FSIL design has located a QTL of interest to a specific chromosomal region, flanking and spanning markers can be defined, so that favourable QTL alleles are defined within specific haplotypes of marker alleles. Such marker–QTL complexes could be treated as quasi-Mendelian super-alleles, and serve as a source of favourable QTL alleles for marker-assisted introgression into the commercial population.

In addition to the above, a FSIL population can be

created as an adjunct to a commercial breeding population, using founder animals from the breeding population itself. In this case, mapping information from the FSIL design population would be directly applicable to the commercial population. This could be used as a basis for linkage disequilibrium mapping or marker-assisted introgression, as above. In addition, multi-generational BLUP-based methods for identification of marker–QTL phase when QTL effect and location are known are under active development (Fernando & Grossman, 1989; Meuwissen & Goddard, 1996). Thus one can envision a programme in which a FSIL design population is established while simultaneously DNA samples and data are accumulated in the commercial breeding population. Mapping information obtained in the FSIL design population would provide a basis for retroactive genotyping of the commercial population leading to BLUP estimates of marker–QTL phase in the chromosomes of the commercial population.

APPENDIX. A note on the Nd^2 method of summarizing power information

In the course of the power analyses, it was shown that power is a function of Nd^2 for all designs, where N is the total size of the mapping population and d is the standardized gene effect. Thus, to approximate power for any combination of N and d within a given mapping design, one need only calculate or simulate power for a small number of Nd^2 values that among them cover a range of power from 0·05 to 0·95. From these, power for all combinations of N and d can then be obtained by calculating the corresponding Nd^2, and interpolation.

For rapid estimates of power with half-sib designs, one can simply calculate the corresponding Nd^2 for an inbred line cross, and solve easily for power with given Type I error. For example, Knott *et al.* (1996, table 2, 10 cM interval) found a power of 0·92 with Type I error 0·001 11 for the case of a half-sib design with 20 sires, 100 progeny per sire; a QTL having two alleles at equal frequency, with effect of d 0·545; a single marker 0·05 cM from the QTL, the marker having four alleles at equal frequency (giving a probability of 0·75 of a sire being heterozygous); and dam marker genotype information not available (giving a probability of 0·75 of determining which sire allele an offspring has inherited). The marker-associated effect of a QTL with effect d, at a distance r cM from the markers, is equivalent to the effect of a QTL with effect $(1-2r)d$ at the marker. Thus, the power obtained corresponds to a value of $Nd^2 = 2000\,(1-0{\cdot}10)^2\,(0{\cdot}545)^2\,(0{\cdot}75)\,(0{\cdot}75) = 270{\cdot}6$, for a completely informative marker, at the QTL. Dividing by an RN of 6 gives the equivalent $N(1-2r)^2d^2 = 45{\cdot}1$ for an inbred line design. Following the well-known expression for the power of an F2 design (Soller *et al.*, 1976), we have

$$N(1-2r)^2d^2 = 2(z_{\alpha/2} + z_{1-\beta})^2.$$

Substituting $N(1-2r)^2d^2 = 45{\cdot}1$ and $z_{\alpha/2=0{\cdot}000\,555} = 3{\cdot}265$, yields $z_{1-\beta} = 1{\cdot}48$, and power $= 0{\cdot}93$. Clearly, this is a very close approximation to the value obtained in the simulation study.

This research was supported by the United States–Israel Binational Science Foundation (BSF), the United States–Israel Binational Agricultural Research and Development Fund (BARD) and the MAGNET Program of the Israel Ministry of Science. We thank two anonymous referees for their constructive comments.

References

Beckmann, J. S. & Soller, M. (1988). Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations. *Theoretical and Applied Genetics* **76**, 228–236.

Darvasi, A. (1997). Interval specific congenic strains (ISCS): an experimental design for mapping a QTL into a one centimorgan interval. *Mammalian Genome* **8**, 163–167.

Darvasi, A. & Soller, M. (1992). Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics* **85**, 353–359.

Darvasi, A. & Soller, M. (1994). Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* **138**, 1365–1373.

Darvasi, A. & Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199–1207.

Fernando, R. L. & Grossman, M. (1989). Marker-assisted selection using best linear unbiased prediction. *Genetics, Selection, Evolution* **21**, 467–477.

Haley, C. S., Knott, S. A. & Elsen, J. M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207.

Knott, S. A. & Haley, C. S. (1992). Maximum-likelihood mapping of quantitative trait loci using full-sib families. *Genetics* **132**, 1211–1222.

Knott, S. A., Elsen, J. M. & Haley, C. S. (1996). Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**, 71–80.

Lande, R. & Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743–756.

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Lipkin, E., Mosig, M. O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A. & Soller, M. (1998). Mapping loci controlling milk protein percentage in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers. *Genetics* **149**, 1557–1567.

Meuwissen, T. H. E. & Goddard, M. E. (1996). The use of marker haplotypes in animal breeding schemes. *Genetics, Selection, Evolution* **28**, 161–176.

Soller, M. (1994). Marker-assisted selection, an overview. *Animal Biotechnology* **5**, 193–208.

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., Brody, T. & Genizi, A. (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* **47**, 35–39.

Van der Beek, S. J., van Arendonk, J. A. M. & Groen, A. F. (1995). Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. *Theoretical and Applied Genetics* **91**, 1115–1124.

Weller, J. I., Kashi, Y. & Soller, M. (1990). Daughter and granddaughter designs for mapping of quantitative trait loci in dairy cattle. *Journal of Dairy Sciences* **73**, 2525–2537.