Power of a chromosomal test to detect genetic variation using genetic markers

PETER M. VISSCHER*† & CHRIS S. HALEY‡

†*Institute of Ecology and Resource Management, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, U.K. and ‡Roslin Institute, Roslin, Midlothian EH25 9PS, U.K.*

The power of a backcross design to detect genetic variation associated with a single chromosome was investigated. A simple chromosomal test was suggested in which the phenotypic observations are regressed onto genotypic information from multiple markers. It was shown that the optimum marker spacing depends on the underlying genetic structure and chromosome length. A sparse marker map, with markers approximately every 50 cM, is sufficient to detect chromosomal variation if the nature of the genetic variance is coupled polygenes, whereas the optimum marker spacing to detect a single QTL somewhere on the chromosome is slightly denser, about 20–40 cM. Although the method was demonstrated for line crosses, it can equally be applied to other populations, for example four-way crosses and half-sib designs.

Keywords: chromosomal test, genetic markers, inbred lines, interval mapping, QTL, regression.

Introduction

With the advent of molecular marker technology, sophisticated statistical methods have been developed to map quantitative trait loci (QTLs) in experimental and commercial populations of plants and animals (e.g. Lander & Botstein, 1989; Haley & Knott, 1992; Jansen, 1993, 1994; Zeng, 1993, 1994; Haley *et al.*, 1994; Knott *et al.*, 1997). These methods have in common that they focus on mapping a single or multiple QTLs of moderate to large effect on a chromosome. However, the actual genetic architecture in a population under study is likely to be a distribution of QTL effects, with some individual QTL effects large enough to be detected, and many effects so small that they contribute to polygenic variation. For a genome-wide scan, few general strategies have been suggested to elucidate the QTL configuration across the genome. One of the most general methods is the MQM method of Jansen (1993, 1994), which consists of first identifying regions on different chromosomes which explain phenotypic variation by regression onto individual markers. This is followed by the mapping of multiple QTLs by applying interval mapping whilst fitting significant markers from other regions as cofactors

in the model. Zeng (1993, 1994) proposed an almost identical method.

It can be argued that a natural strategy to map QTLs in a genome-wide scan is to start by identifying those chromosomes which explain a significant proportion of variation, and then to dissect the QTL configuration on those chromosomes by mapping single or multiple QTLs (Visscher & Haley, 1996). This approach potentially allows the use of a relatively low resolution genome scan in the first instance, followed by additional genotyping in regions that contain significant genetic variation in order to dissect its causes in more detail. To detect genetic variation on a single chromosome, a chromosomal test has been suggested, in which phenotypes are regressed on a number of markers on that chromosome (Visscher & Haley, 1996). The rationale of such a test is that those chromosomes that do not explain a significant amount of variation should be excluded from further analyses. An additional attraction of the approach is that it simplifies the problem of setting significance thresholds (e.g. Lander & Kruglyak, 1995), because in the first instance the number of independent tests equals the number of chromosomes. This test has been applied in data analysis in trees (Knott *et al.*, 1997), pigs (Knott *et al.*, 1998) and dairy cattle (De Koning *et al.*, 1998).

*Correspondence. E-mail: peter.visscher@ed.ac.uk

Depending on the amount of genetic variance, and its exact nature (single QTL, multiple QTLs, polygenic), there will be an optimum marker spacing to be used for the chromosomal test. For example, if too few markers are used, there is a chance that the genetic variance may not be detected. If too many markers are used, the high correlation between adjacent markers means that markers are being included that explain little variation, so that the test fails to be significant. In this study we explore the factors that determine the optimum marker spacing, and give some empirical results for the power of a chromosomal test under various QTL configurations. We compare the power of the chromosomal test with the power of interval mapping using a very dense marker map.

Methods

We use a backcross (BC) or F_2 population of size N, with fully informative markers. The amount of genetic variation on a single chromosome of interest, as a proportion of the phenotypic variance, is h_c^2. The test to detect genetic variance on the chromosome is based on fitting m markers, and the test statistic is distributed as an F-test under the null hypothesis of no genetic variation associated with that chromosome. For large N, the test statistic is approximately distributed as a χ^2. For a backcross population, the F-test has $\{m, N-m-f\}$ degrees of freedom and the χ^2 test has m degrees of freedom, where f is the number of additional fixed effects, including the mean, in the model. For an F_2 population, the degrees of freedom are $\{2m, N-2m-f\}$ and $2m$, respectively, because it is assumed that two parameters (two degrees of freedom) are fitted per marker. In the presence of genetic variance, the test statistic is distributed as noncentral F, or, approximately, as a noncentral χ^2. For a single marker coincident with a single QTL, the noncentrality parameter is $N\alpha^2/2$ or $N\alpha^2/4$, for F_2 and BC populations, respectively, where $\{2\alpha\}$ is the difference between the parental lines for the QTL alleles. If the markers fitted fully explain the genetic variance (var (A)), then the noncentrality parameter is $N\{$var $(A)/$var $(E)\} = Nh_c^2/(1-h_c^2)$, with var (E) the environmental variance. However, because the markers' positions are unlikely to coincide with QTLs on that chromosome, the markers only explain a proportion of the variance. Depending on the underlying genetic structure, the proportion of variance explained by the markers can be determined.

Allowing for the fact that the markers do not account for all of the genetic variation on the chromosome, the noncentrality parameter becomes:

$$\lambda = N R_{\{L,m\}}^2 h_c^2/(1-h_c^2), \tag{1}$$

with $R_{\{L,m\}}^2$ the amount of genetic variance explained by m markers on a chromosome of length L.

The power of a chromosomal test is calculated as the probability that the test statistic exceeds a given threshold value:

$$\text{Power} \mid (N, h_c^2, m, L) = \text{Prob } (T > \text{THRES}_\alpha), \tag{2}$$

with T the test statistic (which follows a noncentral χ^2 or F), and THRES_α the threshold of a central χ^2 or F pertaining to a Type I error of α. The thresholds for the power calculations were calculated using a central χ^2 distribution, on the basis that N will usually be large enough to warrant this approximation. Given the values of N, h_c^2, m and L, the noncentrality parameter was calculated according to eqn (1), and the power was determined using an approximation to the noncentral χ^2 distribution (Abramowitz & Stegun, 1964).

Coupled polygenes

We define the coupled polygenic model as a large number of QTLs in coupling, with small effect spread evenly throughout each chromosome (Visscher & Haley, 1996), where the amount of genetic variance explained by the markers was determined previously for BC populations and four-way crosses (Visscher, 1996; Knott et al., 1997). In the case of coupled polygenes, the amount of genetic variance is proportional to the variation in genomic proportion (e.g. Hill, 1993). Genomic proportion is defined as the proportion of the genome which originates from either of the two founder lines. Because it is assumed that those founder populations are fixed for alternative alleles at many linked loci each with the same effect on the trait (Visscher & Haley, 1996), it follows that the variation in genomic proportion is proportional to the genetic variance attributable to the many linked loci. For F_2 populations, the same R^2 can be used because all the relevant components, i.e. the variance in marker scores, the variance in genomic proportion and their covariance, are scaled by a factor of 2 relative to the BC population.

Single QTL

In the case of fully informative markers, the proportion of QTL variation explained by the markers is a

function of the distance between the QTL and the nearest markers. If the QTL is located outside a marker bracket (i.e. to the left of the leftmost marker, or to the right of the rightmost marker on the chromosome), the proportion explained, assuming Haldane's mapping function, is:

$$R^2_{\{M\}} = (1 - 2r_{\text{QTL},M})^2 = \exp\{-4d\}, \tag{3}$$

with $r_{\text{QTL},M}$ the recombination rate between the QTL and the nearest marker M, and d the corresponding distance in Morgans. If the QTL is flanked by two fully informative markers, $M1$ and $M2$, the proportion of variance explained is:

$$\begin{aligned} R^2_{\{M1,M2\}} &= [(1-2r_1)^2 + (1-2r_2)^2 - 2(1-2r_1)(1-2r_2) \\ &\quad (1-2r_m)]/[1-(1-2r_m)^2] \\ &= [\exp\{-2d_1\} + \exp\{-2d_2\} - 2\exp\{-4\,d_m\}]/ \\ &\quad [1-\exp\{-4\,d_m\}], \end{aligned} \tag{4}$$

with r_1 and r_2 the recombination rates between the QTL and the first and second marker, respectively, and d_1 and d_2 the corresponding distances in Morgans. The recombination rate between the flanking markers is r_m, with corresponding distance d_m. These results follow directly from the calculations of Haley & Knott (1992) and Whittaker *et al.* (1996).

For a given location of a single QTL on the chromosome, the R^2 value can be calculated for any given marker locations, given eqns (3) and (4). Alternatively, the average power of detecting variation from a single QTL can be calculated assuming a uniform distribution of the location of that QTL. That is, assuming that the QTL is located at 0, 1, 2, ..., L cM, the R^2 value and subsequently the power can be calculated for each possible QTL position, and the power corresponding to a QTL of 'average location' calculated by averaging the powers of all $L+1$ positions. In addition to the average power calculated in this way, an element of risk can be calculated by determining the standard deviation of power for all positions along the chromosome.

Markers and heritability

It was shown previously that, in the case of equally spaced markers, the maximum variance in genomic proportion is explained when the two outermost markers are positioned slightly in from the ends of the chromosome (Visscher, 1996). For example, for a chromosome of 100 cM, the optimum positions of two markers to obtain the maximum R^2 were 27 and 73 cM. However, the loss in precision of estimating the amount of genomic proportion is small when an equal marker spacing is used which treats the

chromosome as if it were circular, that is, assuming a marker spacing between the two distal markers which is the sum of the distance between the markers and the chromosome ends. Therefore, the following marker spacing was used throughout this study: (i) in the case of a single marker, it was positioned in the middle of the chromosome; (ii) with multiple markers, the distance between the markers was $\Delta = L/m$, and the first marker was placed at position $\Delta/2$. For example, for two markers on a 100 cM chromosome, $\Delta = 50$ cM and the marker positions are 25 and 75 cM.

The power of the chromosomal test was calculated as a function of $x = Nh^2_c/(1-h^2_c)$. For a typical genome of 2000 cM, the average heritability per 100 cM is ≈ 0.02 for a trait with a heritability of 0.40 in a BC or F_2 population. Therefore, the values of x taken into account were in the range 1–20 for $L = 100$, and 10–100 for $L = 500$. A value of $x = 1$ corresponds, for example, to the case of $N = 100$ and $h^2_c = 0.01$, or $N = 500$ and $h^2_c = 0.002$, and $x = 20$ corresponds to $N = 100$ and $h^2_c = 0.167$, or $N = 500$ and $h^2_c = 0.038$.

Interval mapping

One obvious alternative approach to using the chromosomal test is to apply interval mapping (Lander & Botstein, 1989; Haley & Knott, 1992) along the chromosome. Depending on the actual genetic architecture, interval mapping may be more or less powerful than other mapping strategies. Powers were calculated for interval mapping assuming a single QTL randomly positioned along the chromosome. For interval mapping, it was assumed that a dense marker map was used, so that the R^2 value was always unity. The corresponding threshold was calculated from the Lander & Botstein (1989) infinitely dense map approximation. For a single chromosome with $L = 100$ and $L = 500$, these thresholds for the F-ratio are 9.1 and 12.5, respectively. Powers were compared to that of the chromosomal test, fitting a range of numbers of markers for the latter.

Results

The power to detect genetic variation was calculated for coupled polygenes and for variance resulting from a single QTL. In the case of a single QTL, both the power corresponding to a fixed position of the QTL, and the average power assuming a uniform distribution of the location of the QTL on the chromosome were calculated. The length of the

chromosome was either 100 or 500 cM, and the number of equidistant markers varied from 1 to 20.

In Table 1, the variance in genomic proportion that is explained by the markers is shown, for $L = 100$ and $L = 500$. The R^2 for both the optimum marker spacing, determined by a search algorithm (see Visscher, 1996, for details), and the approximate marker spacing actually used are shown. The results for $L = 100$ cM are the same as in Visscher (1996). Clearly, the approximation is very reasonable. For $L = 100$ cM, three markers are sufficient to detect > 0.9 of the variation in genomic proportion, whereas for $L = 500$ cM, more than nine markers are needed to achieve $R^2 > 0.9$.

Coupled polygenes

For $L = 100$ cM, the power of a chromosomal test to detect coupled polygenic variation is shown in Fig. 1. It appears that one or two markers per chromosome are sufficient, and that fitting more than two markers reduces power. The curves for the different marker spacings hardly overlap, so that the same marker spacing (one or two markers) has the best power, regardless of population size or heritability. Hence, the optimum marker spacing is about 50 cM. However, the loss in power by fitting markers every 20 cM (i.e. results for five markers) is generally less than 10%.

For a much longer chromosome, $L = 500$ cM, the power curves plotted in Fig. 2 for a range of $Nh_c^2/(1-h_c^2)$ of 10–100 are very different from those

of the shorter chromosome, in that the overall power is lower for the overlap of $Nh_c^2/(1-h_c^2)$ with Fig. 1 (range 10–20), and the optimum number of markers differs. The maximum power is achieved by fitting 10–20 markers. Hence, the optimum marker spacing to detect coupled polygenic variance for a long chromosome is in the range 25–50 cM. When more than 20 markers were fitted, power decreased because the additional parameters fitted did not explain more variation (results not shown). In practice, fitting too many markers might lead to estimation (colinearity) problems if there are no recombinations between one or more pairs of markers.

Single QTL

The power to detect genetic variation that is caused by a single QTL at positions 0, 25 and 50 cM, respectively, was calculated for a 100 cM chromosome. It was found that the power and optimum marker spacing are extremely dependent on the location of the QTL. For example, when the QTL is outside the location of the distal markers (Fig. 3), the lowest power is achieved by fitting a single marker which is at 50 cM, because the $R^2 = \exp\{-4 \times 0.50\} = 0.135$ (eqn 3), which is extremely low. The optimum marker spacing for a QTL at 0 cM is between 10 and 20 cM. For a QTL positioned at 25 cM, the obvious optimum number of markers is two, because one of the markers then coincides with the location of the QTL (results not

Table 1 Proportion of variance in genomic proportion explained by markers, for a chromosome of length 100 or 500 cM

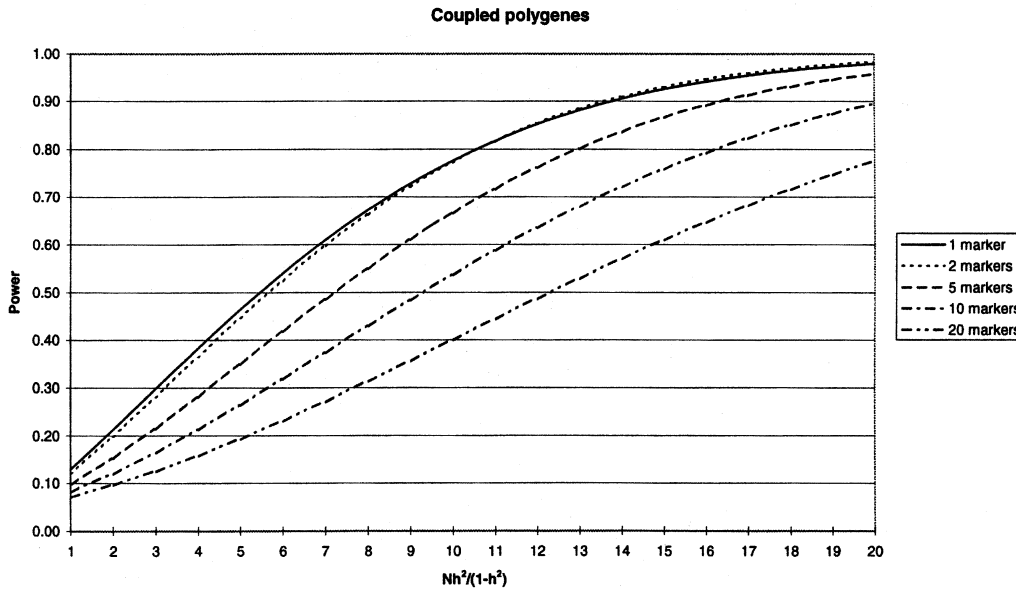| Number of markers	Chromosome length			
	100 cM Marker spacing		500 cM Marker spacing	
	Optimum	Approximate	Optimum	Approximate
1	0.704	0.704	0.219	0.219
2	0.884	0.882	0.415	0.406
3	0.942	0.941	0.568	0.556
4	0.966	0.966	0.678	0.670
5	0.978	0.978	0.757	0.751
6	0.984	0.984	0.812	0.808
7	0.988	0.988	0.851	0.849
8	0.991	0.991	0.880	0.879
9	0.993	0.993	0.902	0.901
10	0.994	0.994	0.918	0.918
20	0.998	0.998	0.978	0.978

Coupled polygenes

Fig. 1 Power of a chromosomal test to detect variance resulting from coupled polygenes depending on the population size, heritability and marker density for a BC population. Chromosome length is 100 cM.

shown elsewhere). All other marker spacings are inferior, with both the widest (single marker) and narrowest (20 markers) spacings giving poor power. For a single QTL at location 50 cM, the optimum number of markers is one, because the marker is placed at the same location as the QTL. Fitting any odd number of markers gives an R^2 of unity, because the middle marker always coincides with the QTL

position. Hence, three (not shown) and five markers are superior over other marker densities.

Average QTL position

Results for the average power pertaining to a single QTL that is uniformly distributed along the chromosome are shown in Figs 4 and 5. For the 100 cM
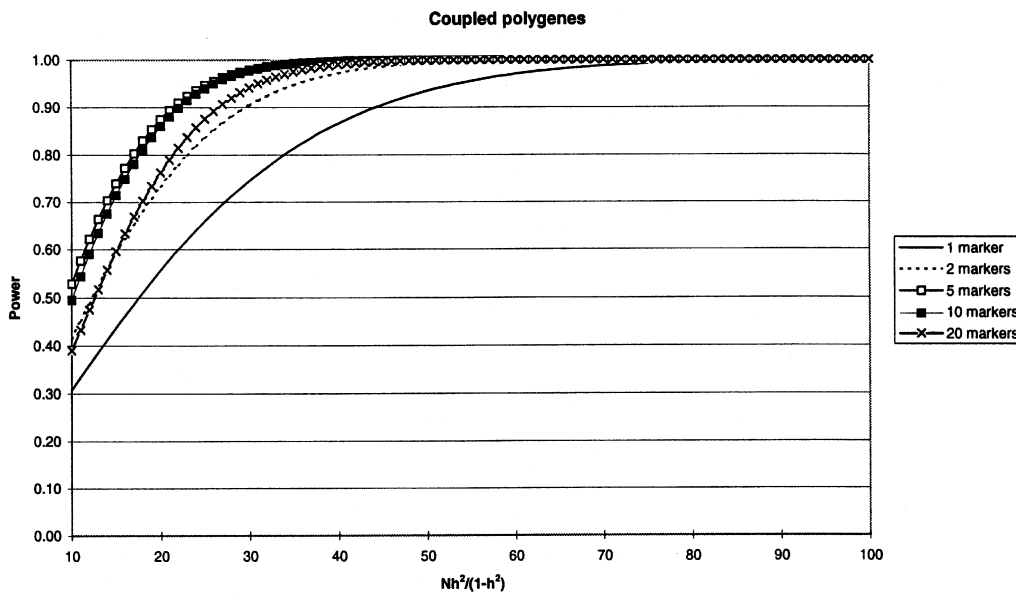
Coupled polygenes

Fig. 2 Power of a chromosomal test to detect variance resulting from coupled polygenes depending on the population size, heritability and marker density for a BC population. Chromosome length is 500 cM.
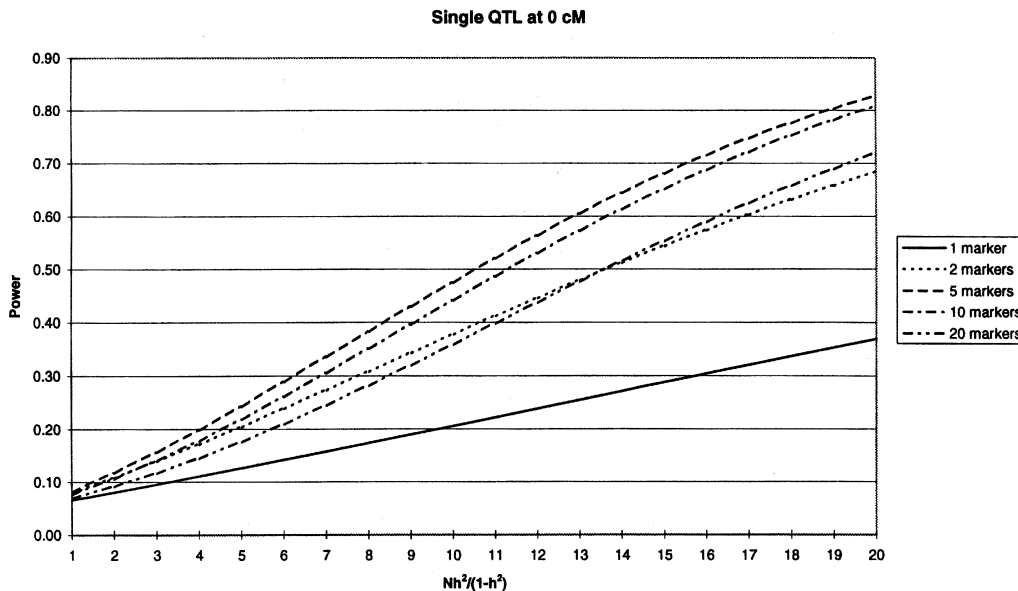
Single QTL at 0 cM

Fig. 3 Power of a chromosomal test to detect variation resulting from a single QTL depending on the population size, heritability and marker density for a BC population. Chromosome length is 100 cM, and the QTL is located at 0 cM.

chromosome (Fig. 4), the average power curves are similar for a range of marker densities. In particular, the curves for two and five markers are similar, which suggests that in practice a marker spacing of about 20–50 cM is optimum. About 10% in power is lost when fitting 10 instead of five markers. For the longer chromosome (Fig. 5), too few markers is clearly inferior. The optimum number of markers is 10–20 markers, corresponding to a marker spacing of 25–50 cM.

Interval mapping

Results comparing the power to detect a single QTL for interval mapping and the chromosomal test are shown in Table 2. For low powers, corresponding to
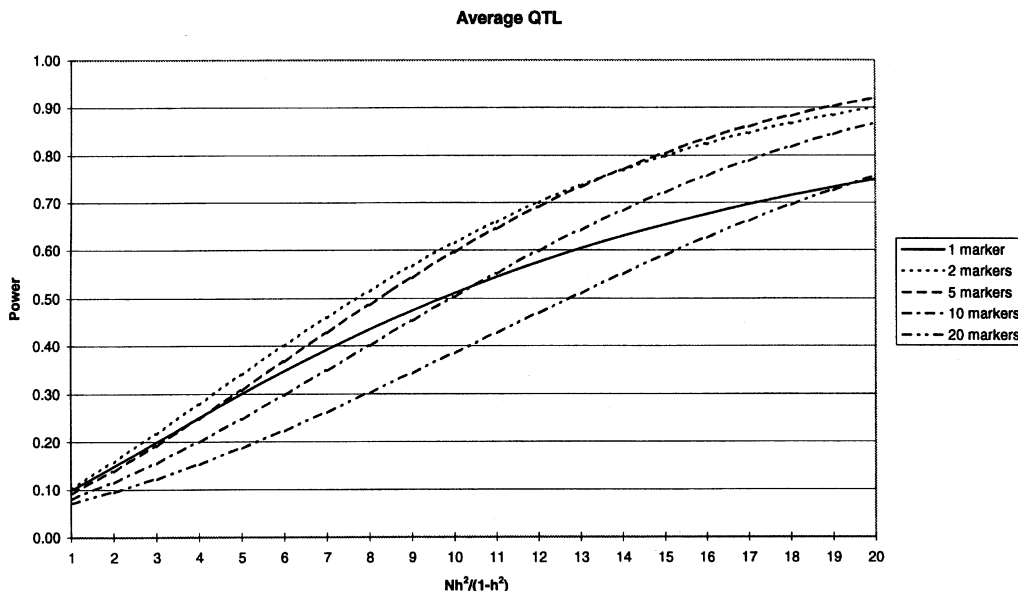
Average QTL

Fig. 4 Average power of a chromosomal test to detect variation resulting from a single QTL depending on the population size, heritability and marker density for a BC population. Chromosome length is 100 cM, and the QTL is uniformly distributed on the chromosome.

Average QTL

Fig. 5 Average power of a chromosomal test to detect variation resulting from a single QTL depending on the population size, heritability and marker density for a BC population. Chromosome length is 500 cM, and the QTL is uniformly distributed on the chromosome.
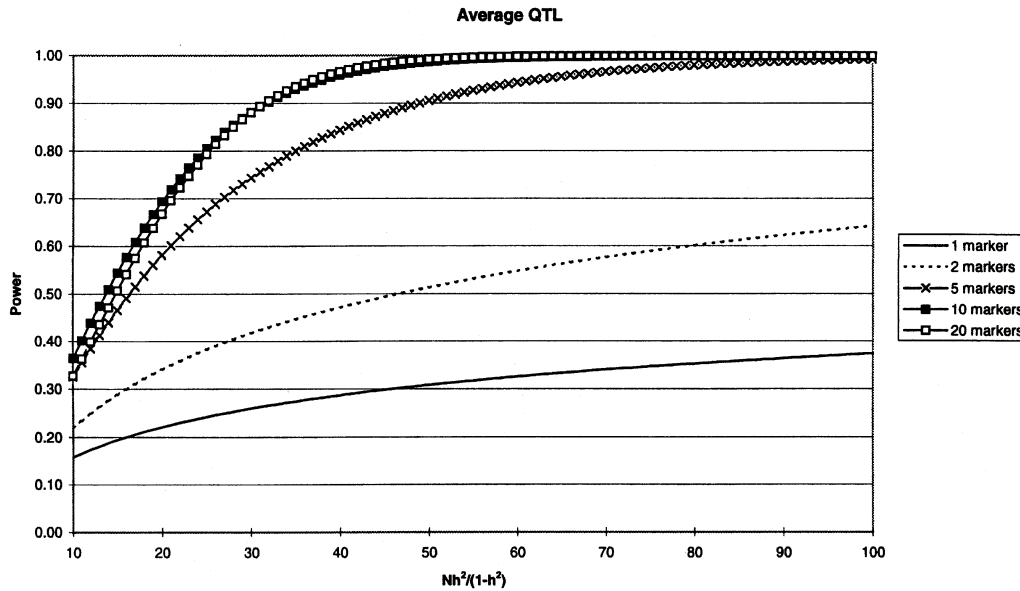
a small amount of variation explained by the single QTL on that chromosome, the chromosomal test is more powerful than interval mapping. For example, the power for the chromosomal test was 30% to detect a heritability of 1% caused by the single QTL when using four markers on a 100 cM chromosome, whereas the corresponding power of interval mapping was 21% (Table 2). However, for intermediate and higher powers on the larger chromosome ($L = 500$), interval mapping is equal to or better than the chromosomal test. For example, the

power to detect a heritability of 3% for the chromosomal test was 56% when fitting 10 markers, and 64% for interval mapping. The variation in the power along the chromosome, as measured by the standard deviation of the power, was 11% in that case. There is no variation in the power for interval mapping, because it was assumed that the relative amount of genetic variation detected by the dense marker map was always 100%.

A final comparison was made in which the power of interval mapping was calculated under coupled

Table 2 Power ($\times 100$) of the chromosomal test and its standard deviation ($\times 100$) for two marker spacings (number of markers $m = 2$, 4), and power ($\times 100$) for an interval mapping test for a QTL randomly positioned along a chromosome of length $L = 100$ and $L = 500$ cM

	L = 100 cM			L = 500 cM		
	Chromosomal test			Chromosomal test		
h_c^2 ($\times 100$)	$m = 2$	$m = 4$	Interval mapping	$m = 10$	$m = 20$	Interval mapping
1	34 (8)	30 (2)	21	19 (4)	16 (1)	9
2	62 (12)	59 (4)	56	37 (8)	33 (3)	35
3	81 (11)	80 (3)	83	56 (11)	52 (5)	64
4	91 (7)	92 (2)	95	72 (11)	69 (5)	85
5	96 (5)	97 (1)	99	83 (9)	82 (5)	95
10	100 (0)	100 (0)	100	99 (1)	100 (0)	100

polygenic inheritance. A problem arises in how to determine the average proportion of polygenic genetic variation that is accounted for when using interval mapping. We assumed that this proportion is approximated by the proportion of coupled polygenic variation that is detected by a single optimally spaced marker. The proportion of genetic variance attributable to coupled polygenes that is detected by a single marker is, on average, largest when that marker is in the middle of the chromosome (Visscher, 1996). Hence, the R^2 values for interval mapping were calculated for a single centrally positioned marker, and the power was calculated assuming a threshold for a dense marker map. Results are shown in Table 3. The power of interval mapping to detect genetic variance is generally low, in particular for the long chromosome. For example, for a polygenic heritability of 0.05 on a chromosome of 500 cM, the power to detect genetic variance using interval mapping is only 12%. The corresponding power for the chromosomal test, using the same R^2 value (i.e. fitting a single marker only), is also shown in Table 3. These powers are much larger, because the chromosomal test explains more of the variance and a lower significance threshold was used to detect significant genetic variance on the chromosome.

Discussion

In this study we have investigated the power of a backcross design to detect genetic variation associated with a single chromosome. A simple chromosomal test was suggested in which the phenotypic observations are regressed onto genotypic information from multiple markers. We have shown that the optimum marker spacing depends on the actual genetic architecture and chromosome length. Although the method was demonstrated for line crosses, it can equally be applied to other populations, for example four-way crosses (Knott *et al.*, 1997) and half-sib designs (De Koning *et al.*, 1998).

The results suggest that a sparse marker map, with markers spaced approximately every 50 cM, is sufficient to detect chromosomal variation if the nature of the genetic variance is coupled polygenes. On average, the optimum marker spacing to detect a single QTL somewhere on the chromosome is slightly denser, about 20–40 cM.

In practice, the objective of genome scans is not just to partition the variation among chromosomes. Usually, the main objective is to identify regions of the genome that cause a significant proportion of the observed variation. If the chromosomal test is used as a first step in a sequence of tests, a relatively sparse map could be used to identify those chromosomes on which more detailed analysis should be performed. This will, however, generally require denser marker information, because it is not possible to dissect the causes of variation with few markers. At the extreme, for example, the optimum density of two markers on a 100 cM chromosome may have been used for the chromosomal test. With only two markers it is not possible to infer the presence of more than a single QTL with least squares analysis (Whittaker *et al.*, 1996) and maximum likelihood analysis would be almost as severely compromised.

An obvious alternative approach and one that is currently widely used is to perform interval mapping throughout the genome. The balance of advantages between the two methods will depend upon genome structure and the underlying genetic structure for the trait. As might be expected, results from Table 2 indicate that interval mapping may be a superior strategy in a large genome with only a single major QTL segregating. For the smaller chromosome with a single QTL, the powers of the two methods are

Table 3 Power ($\times 100$) for the chromosomal test (using a single marker) and interval mapping test under variance resulting from coupled polygenes for a chromosome of length $L = 100$ and $L = 500$ cM

	$L = 100$ cM		$L = 500$ cM	
$h_c^2(\times 100)$	Chromosomal test	Interval mapping	Chromosomal test	Interval mapping
1	47	12	18	1
2	78	35	31	2
3	93	61	45	4
4	98	80	57	8
5	100	91	68	12
10	100	100	96	46

similar and under the assumed infinitesimal model the chromosomal test is generally superior (Table 3). In the latter case the power from interval mapping is not only inferior, but in addition the wrong inference would be made, because the method would locate a single QTL when in fact there are many. The superior power of interval mapping when analysing a long chromosome with only a single QTL is unsurprising, because many redundant markers (i.e. all those that do not flank the QTL) are fitted in the chromosomal test. In fact, it is perhaps surprising how well the chromosomal test does in terms of power relative to interval mapping in this case.

We have deliberately chosen extreme, and to that extent unrealistic, genetic models which provide boundaries for the performance of the two analytical methods. The range of possible underlying genetic structures is unclear. Most QTL mapping studies detect up to around 10 QTLs for any individual trait, which explain a major proportion, but not all, of the genetic variance (e.g. Stuber, 1995). It is also interesting to note that for a sizeable fraction of QTLs detected the high-scoring QTL allele comes from the low-scoring line. The limited power of QTL studies means that QTLs of small effect will not have been detected and estimates of effects of detected QTLs will not be precise and may be inflated. Therefore, these analyses provide only a limited guide to the likely range of underlying genetic situations. However, we can conclude from these studies that a chromosome representing a sizeable proportion of the genome is likely to be carrying more than a single major QTL. In addition, relatively often the effects of linked QTLs are in opposite directions, and interval mapping has been shown to lose substantial power in these circumstances (Haley & Knott, 1992). As soon as there is more than one QTL on a chromosome, the relative power of the chromosomal test compared to interval mapping will increase. Thus, given the relatively good performance of the chromosomal test compared to interval mapping in the extreme case of a single QTL on a large chromosome, the chromosomal test may often have the advantage of power in practice.

There remain some unexplored issues relating to the chromosomal test. In particular, for outbred populations or crosses between outbred lines, markers are not completely informative as they are in the theoretical studies performed here. This means that some markers in some individuals will have missing information (as is usually the case with crosses based on inbred lines because of technical difficulties, etc.). In practice missing markers can be replaced with 'virtual markers' constructed from information from flanking and informative markers. This will, however, change the correlations between adjacent markers from those found when markers are fully informative and this is likely to impact upon the optimum marker densities for the chromosomal test.

The numerical examples in the present studies were all performed for a backcross population. In many livestock QTL mapping experiments, F_2 populations are used, and then a choice can be made in the number of degrees of freedom fitted per marker for the chromosomal test (one or two). For small experimental population sizes, it may be better to fit only a single degree of freedom per marker to avoid losing too many degrees of freedom in the analysis, because for additive effects most variance will be taken out by a single marker effect. In addition, fitting an additive model for multiple markers is consistent with the hypothesis of an infinitesimal coupling model (Visscher & Haley, 1996), whereas a dominance model is not.

We have made some comparisons of the chromosomal test with interval mapping in this paper. It would also be valuable to compare the chromosomal test with MQM mapping (Jansen, 1993, 1994). However, as previously noted, the outcome of any comparisons will very much be dependent on the actual genetic architecture. For example, Jansen (1994) pointed out that the ability to separate linked QTLs of opposite effects strongly depends on the exact locations of the QTLs and on the marker spacing, and that the MQM method would be better than simple interval mapping in these cases. In this respect, the chromosomal test would be no different, in that two closely linked QTLs (distance, say, <20 cM) with opposite effects would not be detected with a very sparse marker map. Analyses of real data sets with the alternative methods would be very helpful to clarify the value of the methods in practice. The work of Knott *et al.* (1998) provides one comparison, and in that study an interval mapping implementation close to MQM mapping gave results similar to a chromosomal test. The study of Knott *et al.* (1998) was, however, undertaken before the calculations on the optimum marker density given here had been performed. There are now plenty of datasets from QTL studies available, and a reanalysis of some of these using alternative methods would be very valuable.

In conclusion, the chromosomal test is easy to apply and has its optimum performance when applied with a low density marker map. Significance

thresholds are readily determined because the number of independent tests equals the number of chromosomes. Unlike interval mapping approaches, the power of the test is maintained across a wide range of models of genetic variation. Once genetic variation has been detected, a denser marker map and a variety of interval mapping or marker regression analysis methods can be used to dissect causation further. The chromosomal test provides a useful additional tool to aid in the analysis of data from QTL mapping studies.

Acknowledgements

References

ABRAMOWITZ, M. AND STEGUN, I. A. 1964. *Handbook of Mathematical Functions*. Applied Mathematical Series, vol. 55. National Bureau of Standards, Washington.

DE KONING, D. J., VISSCHER, P. M., KNOTT, S. A. AND HALEY, C. S. 1998. Strategies for QTL detection in halfsib populations. *Anim. Sci.* in press.

HALEY, C. S. AND KNOTT, S. A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.

HALEY, C. S., KNOTT, S. A. AND ELSEN, J. M. 1994. Mapping quantitative trait loci in crosses between outbred lines using flanking markers. *Genetics*, **136**, 1195–1207.

HILL, W. G. 1993. Variation in genetic composition in backcrossing programs. *J. Hered.*, **84**, 212–213.

JANSEN, R. C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics*, **135**, 205–211.

JANSEN, R. C. 1994. Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, **138**, 871–881.

KNOTT, S. A., NEALE, D. B., SEWELL, M. M. AND HALEY, C. S. 1997. Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. *Theor. Appl. Genet.*, **94**, 810–820.

KNOTT, S. A., MARKLUND, L., HALEY, C. S., ANDERSSON, K., DAVIES, W., ELLEGREN, H. *ET AL.* 1998. Multiple marker mapping of quantitative trait loci in an outbred cross between wild boar and Large White pigs. *Genetics*, **149**, 1069–1080.

LANDER, E. S. AND BOTSTEIN, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 184–199.

LANDER, E. S. AND KRUGLYAK, L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.*, **11**, 241–247.

STUBER, C. W. 1995. Mapping and manipulating quantitative traits in maize. *Trends Genet.*, **11**, 477–481.

VISSCHER, P. M. 1996. Proportion of the variation in genetic composition in backcrossing programs explained by genetic markers. *J. Hered.*, **87**, 136–138.

VISSCHER, P. M. AND HALEY, C. S. 1996. Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. *Theor. Appl. Genet.*, **93**, 691–702.

WHITTAKER, J. C., THOMPSON, R. AND VISSCHER, P. M. 1996. On the mapping of QTL by regression of phenotype on marker-type. *Heredity*, **77**, 23–32.

ZENG, Z.-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 10972–10976.

ZENG, Z.-B. 1994. Precision mapping of quantitative trait loci. *Genetics*, **136**, 1457–1468.