

On the use of double haploids for detecting QTL in outbred populations

VA Martinez^{1,2}, WG Hill¹ and SA Knott¹

¹Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK;

²Departamento de Fomento de la Produccion Animal, Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Casilla 2, Correo 15, Santiago, Chile

The power to detect quantitative trait loci (QTL) using the double haploid (DH), full-sib (FS) and hierarchical (HI) designs implemented in outbred fish populations was assessed for interval mapping using deterministic methods. The predictions were tested using simulation. The DH design was most efficient for the range of designs and parameters considered and was most beneficial when the FS design was not very powerful. The difference between the designs was largest for a low amount of residual genetic variation. Accounting for an increase of the environmental variance due to the genetic constitution of the double haploid

progeny changed the magnitude of the power, but the ranking of the designs remained the same. As large full sib family sizes can be obtained in fish, the practical value of HI designs as a strategy for increasing the power of QTL mapping experiments is limited when compared with the FS design. Overall, the results suggested that the DH design could be a very useful tool for QTL mapping in fish, and of particular importance when the effect of the QTL is low and the residual genetic variation from other chromosomes can be controlled by using multiple markers.

Heredity (2002) 88, 423–431. DOI: 10.1038/sj/hdy/6800073

Keywords: statistical power; QTL mapping; outbred populations; fish; double haploids; molecular markers

Introduction

In recent years genetic marker technology together with new statistical methodology has aided the dissection of complex traits into locus-specific components that explain the observed variation, the so-called quantitative trait loci (QTL) (Geldermann, 1975; Paterson *et al*, 1991; Cheverud *et al*, 1996). In fish, molecular markers have been used primarily in population genetic studies (see for example Bagley *et al*, 1997). However, despite the great economic importance of, for example, several salmon or tilapia species, little is known about the underlying inheritance of traits that may improve production systems.

The great biological flexibility of fish enables the utilisation of different designs that can be implemented relatively easily in practice. Completely homozygous fish can be produced in only one generation using chromosomal set manipulations, without the many generations of inbreeding needed in other vertebrates. These manipulations enable doubling the chromosomes of an haploid gamete (Streisinger *et al*, 1981; Thorgaard, 1992; Corley-Smith *et al*, 1996; Young *et al*, 1996). Gynogenetic double haploid individuals can be obtained by activating the gametes of females with gamma-irradiated sperm, yielding haploid eggs containing only maternal chromosomes. Diploidy is restored using methods that suppress the first

mitotic division (Streisinger *et al*, 1981). Androgenetic double haploid individuals can be produced in a similar way by activating eggs irradiated with gamma rays with normal sperm (Corley-Smith *et al*, 1996; Young *et al*, 1996). Isogenic lines (clonal lines) could be obtained by reproducing these double haploid individuals using the same techniques as described previously (in which case, completely homozygous clonal lines could be obtained from a single individual) or through crosses with sex reversed double haploid individuals (Young *et al*, 1996). Double haploid lines derived from F_1 lines of this form have been utilised to perform QTL analysis for embryonic development rate in rainbow trout (Robison *et al*, 2000). This may not be the optimal use of double haploid lines, however, since forming the clonal lines is expensive and time consuming in species with a long generation interval. This is a particular problem when the objective is to implement these procedures for QTL mapping in a practical breeding programme.

The use of double haploid individuals for QTL analysis has been studied in the context of selfing populations of plants (Jensen, 1989; Knapp *et al*, 1990). The basic framework comprises the utilisation of gametes from F_1 individuals that are chemically treated to double the chromosome number (Luo and Kearsey, 1990; Lynch and Walsh, 1998). In this type of population, all F_1 parents have identical genotypes, with the same linkage phase, and so all the double haploid individuals are completely informative and linkage disequilibrium is maximal (Lynch and Walsh, 1998).

In this paper we investigate the use of chromosomal manipulations to obtain double haploid individuals from

Correspondence: VA Martinez, Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK. E-mail: victor.martinez@ed.ac.uk

Received 29 January 2001; accepted 22 December 2001

parents sampled directly from an outbreeding population. The mapping population can be obtained in a relatively short period of time, without the need to first produce the clonal lines and the F_1 population. Unlike the double haploids from inbred lines, the use of double haploids applied in outbreeding populations would yield families with a variable amount of information about the linkage of a QTL and markers. To be completely informative, a parent has to be heterozygous at both the markers and the QTL. In segregating populations not all parents will be completely informative, and the phase of linkage between the favourable QTL allele and the marker allele will differ across families. Therefore, the power may differ considerably when moving from the double haploids obtained from inbred line crosses to those from outbreeding populations.

The principal aim of this study is to assess the power to detect QTL for different designs that can be implemented within outbred populations, comparing in particular the double haploid (DH) design with more standard designs using normal reproduction and a collection of full and/or half sibs.

Theory

Power Prediction

For the purpose of power prediction the approach used by Weller *et al* (1990) was extended for use with interval mapping, assuming completely informative markers. The test statistic is based on the squared contrast between offspring receiving different marker haplotypes from the parents (van der Beek *et al*, 1995). The QTL was assumed to be in the middle of the marker bracket. For parents informative for the markers, their progeny can be classified according to the presence or absence of recombination between the markers in the bracket. When the QTL is in the middle of the interval most of the information about the presence of a QTL can be obtained from the contrast between the two non-recombinant classes (contrasts between recombinant and non-recombinant classes will contribute little information for detecting linkage between a marker and a QTL. The test statistic (TS) used in this study for detecting a QTL between a marker bracket is:

$$TS = \sum_{i=1}^p C^2 / SE^2(C) \quad (1)$$

where, $\sum_{i=1}^p$ denotes summation over parents, C is the value of the marker contrast between non-recombinant haplotype classes from the same parent and $SE(C)$ is its standard error. Under the null hypothesis (of no QTL segregating within the interval) the test statistic was assumed to follow a central χ^2 distribution with degrees of freedom equal to the number of parents heterozygous for the markers. This assumes that the phenotypic variance is known without error and that the sample size is large. When the variance is calculated from the sample the test statistic (when summed over families) will follow an F distribution (Geldermann, 1975; Weller *et al*, 1990; van der Beek *et al*, 1995). The assumption that the test statistic follows a χ^2 distribution appears to be robust, even when the sample size is small (data not shown).

In order to compute statistical power, the distribution of the test statistic under the alternative hypothesis is

required. In this case the variable is expected to follow a non-central χ^2 distribution, whose non-centrality depends on the number of parents heterozygous at the QTL, the expected value of the marker contrast and its standard error (Weller, *et al*, 1990; van der Beek, 1995). The non-central parameter of this distribution is $NC(h)$ with h being the number of parents heterozygous for the QTL and the degrees of freedom equal to the total number of parents (p). The power is computed as the probability that this non-central χ^2 variable exceeds a critical value given that h out of p parents are heterozygous for the QTL summed over parents. The number of heterozygous parents follows a binomial distribution, with h out of p independent trials (the parents sampled) and with probability of success equal to the expected proportion of heterozygous individuals assuming random matings. Overall the power was calculated as:

$$Power = \sum_{h=0}^p P(h)P(\chi^2[NC(h),p] \geq T) \quad (2)$$

where, $\sum_{h=0}^p$ denotes summation over parents, $P(h)$ is the binomial probability that h out of p parents are heterozygous for the QTL, and $P(\chi^2[NC(h),p] \geq T)$ is the probability that the non-central χ^2 variable is greater than a significance threshold (T) under the null hypothesis (obtained from a central χ^2 distribution, with p degrees of freedom). The non-central parameters needed to perform power calculations for the different designs considered in the present paper are given in detail below.

Designs for QTL mapping experiments in outbred populations

Three different, balanced, two-generation population structures were considered in the present study. In all cases, the mapping population was obtained from completely informative parents for the markers; the individuals sampled to produce the progeny generation were assumed to be unrelated and were genotyped at the markers and the progeny were genotyped and their phenotypes recorded.

An underlying additive model was assumed with a diallelic QTL segregating, with a difference of $2a$ between homozygotes, and a residual genetic component under infinitesimal assumptions unlinked with the QTL (g). The genotypic frequencies in the parental population follow Hardy-Weinberg expectations with allele frequency q for the increasing allele. Interval mapping was assumed, such that a marker bracket of a certain length flanks the QTL in the interval. No interference in recombination events was assumed (ie, the Haldane mapping function was used). Since the population is assumed to be in linkage equilibrium between markers and QTL, the linkage information must be inferred from the contrasts within families. It was assumed further that the variance contributed by the QTL is small compared to the phenotypic variance.

The double haploid (DH) design

Under the double haploid design, dams that generate the progeny generation are randomly sampled from a base population. The diploid state is restored artificially using treatments that enable the genetic contribution of a single gamete to be doubled. The model for the DH design is:

$$y_{ijk} = \mu + d_i + dm_{ij} + e_{ijk} \quad (3)$$

where y_{ijk} is the phenotype for the quantitative trait, d_i is the random effect of the i th dam, dm_{ij} is the fixed effect of the j th marker haplotype classes within the i th dam and e_{ijk} is the residual term.

When double haploid techniques are used to produce the next generation, the progeny population will no longer follow Hardy-Weinberg expectations, there being no heterozygotes and the frequencies of homozygotes being equal to the allele frequencies in the parental population. The expected performance in the progeny generation under any mode of inheritance in the base population is dependent only on the allele frequencies and gene effect and has expectation equal to $a(2q - 1)$.

The genetic variance in the double haploid progeny is, assuming linkage equilibrium, twice the additive variance of the parental generation:

$$Var(QTL(total)) = 4q(1 - q)a^2, \quad (4)$$

and is equally distributed between and within families.

If the dam is heterozygous for the QTL and for the markers, the expected values for alternative haplotypes based on flanking markers are a mixture of the effect of the homozygous QTL genotypes, weighted by the conditional probabilities of the QTL genotype given each marker class. These probabilities and the conditional expectations are presented in Table 1. The expectations under interference were presented previously by Knapp *et al* (1990).

The variance of the marker means depends on the number of progeny belonging to the corresponding marker allele group and the within family variance. Using interval mapping, the expected number of progeny belonging to each marker haplotype class (m) is dependent on the recombination fraction between the two markers (r_{ab}), ie $(1 - r_{ab})n/2$ and $r_{ab}n/2$ with n progeny per family, for the non-recombinant and recombinant classes, respectively. The observed variance amongst non-recombinant individuals for a completely informative dam can be expressed (ignoring small contributions from the QTL) as:

$$VAR(A_i - B_i) \approx E\left(\frac{2(\sigma_g^2 + \sigma_E^2)}{n(1 - r_{ab})}\right) \approx E\left(\frac{\sigma_w^2}{m}\right) \quad (5)$$

where $A_i - B_i$ denotes the marker haplotype (see Table 1) for the non-recombinant class, σ_g^2 is the variance of the residual polygenic component and σ_E^2 is the environmental variance. The expected value of the ratio σ_w^2/m was approximated using a first order Taylor series as:

$$E\left(\frac{\sigma_w^2}{m}\right) \approx \frac{E(\sigma_w^2)}{E(m)}$$

The marker mean variance would also include any variation due to segregation of the QTL within the class, but this is expected to be very small since only double recombinant progeny would contribute and the number of these in non-recombinant classes is expected to be low. Compared with the residual variance, the increase due to segregation of the QTL is negligible and is therefore ignored in equation 5.

Hence, the expectation of the contrast of the non-recombinant marker classes conditional on the individual being heterozygous for the QTL is given by:

$$E[C - DH] = \frac{2a(1 - r_a - r_b)}{1 - r_{ab}} \quad (6)$$

where r_a and r_b denote the recombination fraction between the left and the right marker with the QTL, respectively, and with variance equal to:

$$SE^2(C - DH) \approx \frac{4(\sigma_g^2 + \sigma_E^2)}{n(1 - r_{ab})} \quad (7)$$

Therefore the non-central parameter for the DH design is:

$$NC_{DH}(h) \approx h \frac{E^2(C - DH)}{SE^2(C - DH)} \approx h \frac{na^2(1 - r_a - r_b)^2}{(1 - r_{ab})(\sigma_g^2 + \sigma_E^2)} \quad (8)$$

where h is the number of dams heterozygous for the QTL.

The full sib (FS) design

Pairs of parents (sires (s) and dams (d)) were chosen at random and then mated to produce a number of full-sib families in the progeny generation (with n progeny per family). The model of analysis is:

$$y_{ijk} = \mu + p_i + h_{ij} + e_{ijk} \quad (9)$$

where p_i is the random family effect, h_{ij} is the fixed effect of the j th marker haplotype within the i th family and e_{ijk} is the residual term.

For completely informative markers, there are two possible non-recombinant haplotype contrasts, one from each of the parents that produce the full-sib family (van der Beek *et al*, 1995). For an additive genetic model, the expected value of the contrast for each parent is equal to:

$$E(C - FS) = \frac{a(1 - r_a - r_b)}{(1 - r_{ab})} \quad (10)$$

The variance of the contrast was obtained from the

Table 1 Joint, marginal and conditional probabilities of QTL alleles and genotypes given the genotypes at the flanking markers and conditional expectations given flanking markers for the DH design

Marker haplotype	Probability of gametes carrying the QTL allele Q_1 given marker haplotypes	Frequency of marker haplotypes	Conditional probabilities of QTL genotype Q_1, Q_1 given flanking markers	Conditional expectation given flanking markers
A_1-B_1	$P(A_i, Q_1, B_i)$	$P(A_1-B_1)$	$P(Q_1 Q_1 A_1-B_1, A_1-B_1)$	$E[y_{ij} A_1-B_1]$
A_1-B_1	$(1 - r_a - r_b + r_a r_b)/2$	$(1 - r_{ab})/2$	$(1 - r_a - r_b + r_a r_b)/(1 - r_{ab})$	$a(1 - r_a - r_b)/(1 - r_{ab})$
A_1-B_2	$(r_b - r_a r_b)/2$	$r_{ab}/2$	$(r_b - r_a r_b)/r_{ab}$	$a(r_b - r_a)/(r_{ab})$
A_2-B_1	$(r_a - r_a r_b)/2$	$r_{ab}/2$	$(r_a - r_a r_b)/r_{ab}$	$a(r_a - r_b)/(r_{ab})$
A_2-B_2	$(r_a r_b)/2$	$(1 - r_{ab})/2$	$(r_a r_b)/(1 - r_{ab})$	$-a(1 - r_a - r_b)/(1 - r_{ab})$

The ordered genotype of the dam is $A_1 - Q_1 - B_1 / A_2 - Q_2 - B_2$.

sum of the variances of the expected marker haplotypes under the same assumptions as the DH design. The variance of the marker haplotype mean is the ratio of the within family residual variance ($\sigma_g^2/2 + \sigma_E^2$) to the expected number of progeny belonging to the non-recombinant class:

$$\text{Var}(A_i - B_i) \approx \frac{2(\frac{1}{2}\sigma_g^2 + \sigma_E^2)}{n(1 - r_{ab})} \quad (11)$$

The non-central parameter is:

$$NC_{FS}(h) \approx h \frac{na^2(1 - r_a - r_b)^2}{4(1 - r_{ab})(\frac{1}{2}\sigma_g^2 + \sigma_E^2)} \quad (12)$$

where h is the number of parents (sires and dams) heterozygous for the QTL.

The hierarchical (HI) design

In hierarchical designs, a series of unrelated sires (s) are mated to an independent set of unrelated (d) females. From each dam n progeny are considered in the analysis. The information about linkage is obtained from two different sources. The sire haplotype contrast is obtained from the difference between the paternal half sib progeny inheriting the alternative non-recombinant sire marker haplotypes. The second contrast is obtained from the difference between the full sib progeny that inherit the alternative non-recombinant marker haplotypes from their dam. The model is:

$$y_{ijklm} = \mu + s_i + d_{ij} + sm_{ik} + dm_{ijl} + e_{ijklm} \quad (13)$$

where y_{ijklm} is the progeny observation, s_i is the random effect of the i th sire, d_{ij} is the random effect of the j th dam nested within the i th sire, sm_{ik} is the fixed effect of the sire marker allele, dm_{ijl} is the fixed effect of the dam marker allele and e_{ijklm} is the residual term.

Although the expectations of the contrasts from both sires and dams are the same as the contrast given for the FS design, the variance of the sire haplotype marker mean includes residual variation between dams with a value equal to $\frac{1}{4}\sigma_g^2/d$. However, the haplotype means are no longer independent, due to the cross-classified structure of dams within sires, ie, progeny of the same dam is included in the alternative haplotype marker means, therefore the variance of the contrast includes minus twice the covariance between haplotype marker means. This term is equal to $-2\frac{1}{4}\sigma_g^2/d$. Finally, the variance of the contrast can be approximated as:

$$\text{Var}[C - HI, \text{sire}] \approx \frac{4[\frac{1}{2}\sigma_g^2 + \sigma_E^2]}{dn(1 - r_{ab})} \quad (14)$$

The observed variance of the dam-marker mean is still equal to that expected for the FS design (equation 11).

Under this model with u sires heterozygous at the QTL the non-central parameter for the sire contrast is approximately equal to:

$$NC_{HI,sire}(u) \approx u \frac{dna^2(1 - r_a - r_b)^2}{4(\frac{1}{2}\sigma_g^2 + \sigma_E^2)(1 - r_{ab})} \quad (15)$$

and the non-central parameter for the dam contrast with v dams heterozygous for the QTL is the same as the non-central parameter for the FS design (as in equation 12).

The test statistic for the HI design is distributed as a χ^2 with non central parameter given by the sum of the sire and dam non-central parameters with $s(d + 1)$

degrees of freedom (Lynch and Walsh, 1998; Gomez-Raya and Sehested, 1999). The maternal and paternal contrasts are expected to be independent given the assumption of random matings and that markers are fully informative, ie, we know from which parent each allele has been inherited. The contrasts are expected not to be independent, however, in situations where markers are less informative and the linkage phase needs to be inferred from the marker data. In this case inferences should include information jointly from a linkage group to infer the most likely linkage phase of individuals.

Numerical study

As explained above, the power is computed as the sum over the possible number of heterozygous parents at the QTL of the probability that the test statistic exceeds the threshold value weighted by the probability of observing that number of heterozygous parents (equation 2). The probability $B(p, 2q(1 - q), h)$ of h heterozygous parents out of p was computed from a binomial distribution with parameters p , the total number of parents, and $2q(1 - q)$, the probability that a parent is heterozygous for the QTL under Hardy-Weinberg expectations in the base population. The non-central parameter needed for power calculations was computed using equations 8 and 12, for the DH and FS design, respectively.

The power was computed for the FS and DH as:

$$\text{Power} = \sum_{h=0}^p B(p, 2q(1 - q), h) P(\chi^2[NC(h), p] \geq T) \quad (16)$$

For the HI design the power was computed accounting for the probability that u out of s sires as well as v out of sd dams are heterozygous for the QTL.

$$\text{Power} = \sum_{u=0}^s B(s, 2q(1 - q), u) \sum_{v=0}^{sd} B(sd, 2q(1 - q), v) P(\chi^2[(NC(u)_{HI,sire} + NC(v)_{HI,dam}), s(d + 1)] \geq T) \quad (17)$$

In all the alternatives examined in this study, the environmental variance was held constant ($\sigma_E^2 = 1$), but the residual polygenic variance (σ_g^2) was variable. The heritability of the residual polygenic component ($h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_E^2)$) ranged from 0 to 0.5. Two values of the QTL effect were explored ($\alpha = 0.2$ and 0.4). The probability of Type I errors was fixed at 0.01 in all designs.

For each design, the experimental population (progeny population) size was kept constant (from 500 to 1000 individuals), but the number of full sib families measured was varied, so as to find the optimum number of families to maximise the power for a given population size. For the HI design both the numbers of full sib offspring per female and females mated to each sire were varied. For a fixed full sib family size, different numbers of females mated to each sire were investigated under this design.

Simulation study

The validity of the expressions derived algebraically for power calculations were tested by Monte Carlo simulation under the genetic model outlined previously above. One chromosomal segment of either 10, 20, 30 or 40 cM was flanked by two informative markers with a QTL placed in the middle. The heritability of the residual polygenic component was simulated equal to 0.5. The design of the population was the optimal for each design and

QTL effect according to theoretical results (see below). The total number of individuals genotyped was 1000. The test statistic was computed using only the non-recombinant classes (equation 1). Ten thousand replicates were simulated under the null hypothesis of no QTL segregating. Since the empirical distribution of the test statistic differed little from that obtained theoretically (χ^2 with degrees of freedom equal the number of parents) the theoretical distribution was used to obtain the significance thresholds ($\alpha = 0.01$). One thousand replicates were simulated and the power was computed as the number of replicates that gave a test statistic greater than the theoretical significance threshold.

Results

The method developed in this paper for predicting power appears to be a very good approximation to results obtained using simulations (Figure 1). In all the cases considered, differences between the predicted and the simulated values were lower than 1%, irrespective of the length of the marker bracket utilised and the effect of the QTL.

The variance of the marker haplotype (equation 5) is a first order Taylor series approximation of the expected value of the ratio between the within family variance and the number of non-recombinant progeny for the corresponding marker haplotype. The second order approximation gave results very similar to the first approximation, so the latter was used throughout. Simulations show that equation 5 tends to underestimate the expected value of the ratio, especially for a very low number of progeny (results not shown). Nevertheless, it appears to be a reasonable approximation for the large family size as considered in this paper and representative of the number of progeny available in a single family of fish.

Comparison between the DH and FS design

Effect of population structure: The design of the experiment aimed to detect linkage between a marker and a QTL in an outbred population proved to affect the power markedly. The DH design generally has higher power than the FS and the HI designs and in some of the cases considered, increases of 20% in power were observed.

For a given population size, it appears that the power of the DH design is dependent on the family size that must be reared in the experimental population. This is illustrated in Figure 2 for different QTL effects (0.2 and 0.4) and population sizes (1000 and 500 progeny) when the marker interval is equal to 10 cM. The optimal population structure was dependent on the size of the QTL effects. For the smaller QTL effect considered, the optimal population structure comprised a lower number of families each with a larger number of progeny. For informative designs (a large QTL effect coupled with a large population size), there is a reduction in power if the number of families is reduced (or the number of sibs is increased). With fewer families there is a higher chance of no families being informative for the QTL and the benefit of increasing the number of sibs is very low.

The maximum power when testing one single family equals the binomial probability that the contrast is informative, ie for the DH design that the dam is heterozygous at the QTL (see Figure 2). Greater power is obtained for the FS design when only one family is tested, because each full sib family provides two contrasts instead of one and there is a greater chance that at least one parent is heterozygous for the QTL. For less informative designs, power generally increases when reducing the number of families.

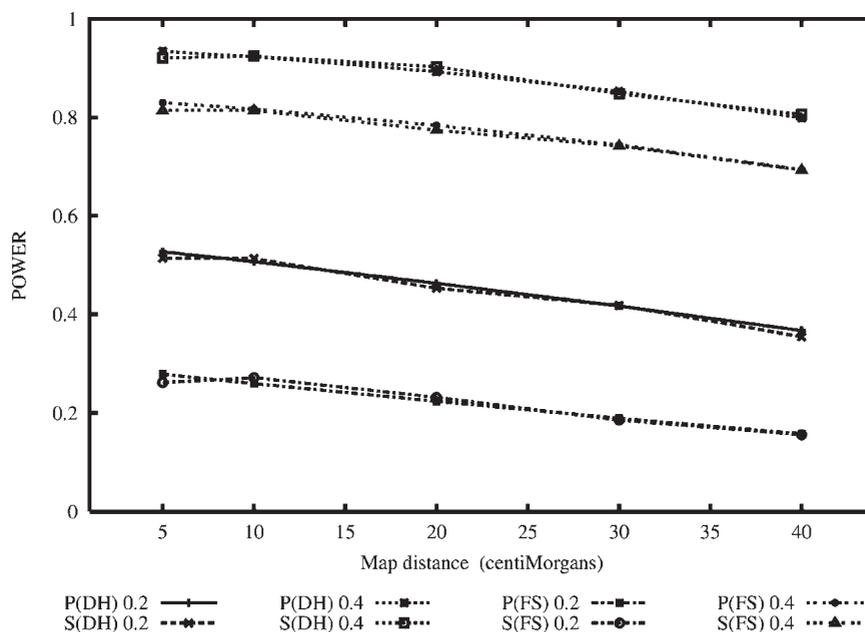


Figure 1 Comparison of power prediction for the optimal double haploid (DH) and full sib (FS) designs using the approximations developed (P) and simulations (S). The QTL effects were 0.2 and 0.4, as explained in the text with a population size equal to 1000 individuals. The residual genetic and environmental variance was set equal to 1. The probability of Type I errors was set to $\alpha = 0.01$.

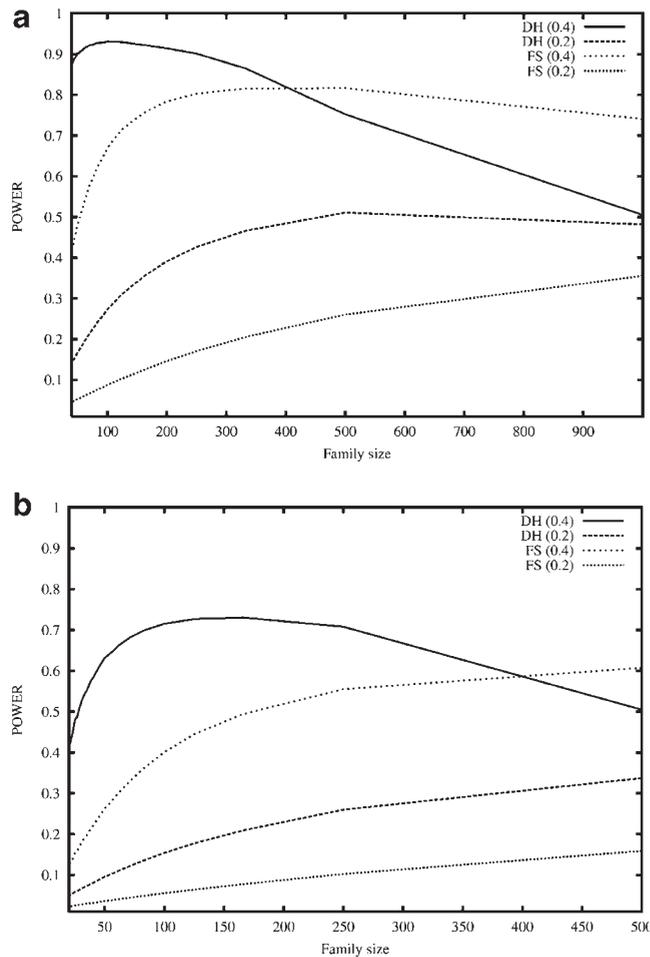


Figure 2 Predicted power as a function of the family size for the double haploid (DH) and full sib (FS) designs, for a constant population size of 1000 (a) and 500 (b) individuals genotyped in the progeny generation. The QTL effect was either 0.2 and 0.4 with equal gene frequencies and both the residual genetic and environmental variance equal to 1. The probability of Type I errors was set to $\alpha = 0.01$.

Expected maximum power calculations

In Table 2, the maximum power obtained as a function of the population size, gene effect and the amount of residual genetic variation is presented. The DH design gave greater power than the FS design in the range of situations considered, the difference being particularly large when the experiment is of small size and the QTL has a small effect (Figure 1).

In all the cases considered, an increase in the size of the population produced an increase in expected maximum power. The overall effect of increasing the total number of individuals measured tended to be larger for the FS design. Increasing the effect of the segregating QTL had similar effect on the power of both designs (see Figure 2a and 2b). Increasing the residual genetic variation has a particularly great effect on the power of the DH design (Table 2), because the genetic variation in the DH progeny is double that expected under normal reproduction (see equation 4). Since some of this background variation due to genes located on other chromosomes can be controlled by using multiple markers in a multiple regression framework (eg, Jansen, 1993; Zeng, 1994), this

may be of less importance when applying the DH design in practice.

Hierarchical (HI) designs appear to have very similar power over the range of mating ratios considered in Table 3, when the family size is relatively high. Predictions are also given for the FS design (a special case of the HI design with a mating ratio 1:1). Increasing the family size further increases the power of the FS over the HI design. For example, for a full sib family size of 200 individuals and a QTL heritability equal to 0.07 ($a = 0.4$, $\sigma_g^2 = 0.1$), power of the FS design is 0.91, compared with 0.83 for the HI design when five females are mated to one male. This is because in this study the information given by the dam contrast in each full sib family is already high due to the large full sib family sizes, therefore using the half sib progeny in addition adds little information. Concomitantly, using a low number of sires increases the chance of sampling individuals that are homozygous for the QTL and, hence, uninformative. In cases where the full sib family size is not sufficiently large, the power of the HI design is increased compared to the FS design (Table 3), irrespective of the magnitude of the residual genetic variation.

Discussion

The objective of this study was to assess the power of QTL mapping under different designs in segregating populations, in particular to compare the efficiency of the use of double haploid families and two generation outbred designs.

The power obtained when utilising offspring only from the non-recombinant marker classes is greater than that obtained when including the less informative recombinant marker classes, even though not all the phenotypic information from a family is used. This result may be explained by noting that the contrast between the non-recombinant classes is an estimator of a multiple of the QTL effect (ie, is twice in the case of the DH design), scaled by a factor $(1 - r_a - r_b)/(1 - r_{ab})$, which for most situations is approximately equal to 1, even when the QTL is in the middle. This contrast is expected to be much larger than all the other contrasts and this benefit far outweighs the disadvantage of using only a subset of the data. The methodology presented here can be used to determine the best alternative among different designs. In practice, all marker classes are included in the analysis using regression or maximum likelihood techniques in order to obtain estimates of QTL effect and location.

The results show that it would be possible to increase significantly the statistical power of QTL mapping experiments by using the DH design, even when compared against the comparatively high power of the FS designs considered here. The DH design would not require a dedicated mapping population, since it is applied directly to the same outbred base population as the FS design. Thus, there is no increase in the time to obtain the progeny generation utilised for mapping. In contrast, the time lag for obtaining the mapping population may be a constraint when using the F_2 design from line crosses (Haley *et al*, 1994) or when developing clonal lines by means of chromosomal manipulations (Robison *et al*, 2000), given the long generation interval in some species of fish. Another advantage of using the outbred population directly rather than creating clonal lines is that,

Table 2 Power (corresponding family size in parenthesis) optimized over family structure for the double haploid (DH) and full sib (FS) designs, for different population size (N), QTL effect (a) and residual genetic variation (σ_g^2). The environmental variance was 1 in all cases

σ_g^2	DH			FS			
	0.0	0.5	1.0	0.0	0.5	1.0	
a	No.	Power			Power		
0.2	200	0.276 (200)	0.180 (200)	0.129 (200)	0.084 (200)	0.065 (200)	0.053 (200)
	400	0.449 (400)	0.355 (400)	0.276 (400)	0.199 (400)	0.151 (400)	0.120 (400)
	600	0.573 (300)	0.449 (600)	0.386 (600)	0.319 (600)	0.248 (600)	0.199 (600)
	800	0.660 (400)	0.531 (400)	0.449 (800)	0.423 (800)	0.341 (800)	0.280 (800)
	1000	0.731 (333)	0.607 (500)	0.507 (500)	0.507 (1000)	0.423 (1000)	0.356 (1000)
0.4	200	0.660 (100)	0.531 (100)	0.449 (200)	0.423 (200)	0.341 (200)	0.280 (200)
	400	0.869 (57)	0.750 (133)	0.660 (200)	0.660 (400)	0.594 (400)	0.531 (400)
	600	0.955 (50)	0.870 (100)	0.787 (150)	0.787 (300)	0.711 (300)	0.660 (600)
	800	0.986 (50)	0.936 (80)	0.869 (133)	0.869 (266)	0.806 (400)	0.749 (400)
	1000	0.996 (50)	0.969 (83)	0.923 (111)	0.923 (250)	0.870 (333)	0.817 (500)

Table 3 Power for the hierarchical and full sib designs given different family sizes, population structures (mating ratio) and residual genetic variation (σ_g^2) for a QTL effect equal to 0.4. The environmental variance was equal to 1 and the population size was 1000

Mating ratio	Family size	σ_g^2		
		0.0	0.10	0.20
1:5	10	0.43	0.39	0.36
1:2	10	0.36	0.33	0.30
1:1	10	0.28	0.25	0.23
1:5	100	0.88	0.86	0.85
1:2	100	0.90	0.89	0.87
1:1	100	0.89	0.87	0.85

depending on the number of parents used, the design can gain access to multiple QTL alleles present in the initial base population.

It was assumed for the purposes of power prediction that completely informative markers were available. In practice, depending on the type of marker used, the levels of heterozygosity will be lower and this will lead to a reduction in power (Knott and Haley, 1992; van der Beek *et al*, 1995). In practice, information for many linked loci is used simultaneously with the aid of multipoint methods, and this tends to ameliorate the decrease in informativeness of only one single interval.

It is not unrealistic to assume the large full sib families considered here for fish, which have a much higher reproductive capacity than other farmed species. Thus higher power can be obtained than for other species in which the number of full sibs per family is a limitation (Knott and Haley, 1992). For example, in Figure 2 it is shown that power decreases rapidly as the full sib family size decreases, even when the total sample size is relatively high.

In the range of situations considered, it appears that maximum power increases when either the population size and/or the size of the QTL increases, as expected. For a constant population size, the power depends upon the number of families. For experiments of low power,

eg for a QTL of small effect and a relatively low population size, power generally decreases as the number of families in the population increases, whereas for experiments of relatively high power, the magnitude of the expected power tends to increase with the number of families in the population. This is consistent with results of Muranty (1996). These results may be explained by the fact that the parents sampled have unknown *a priori* information about the QTL genotype. Under these conditions, the QTL alleles are more accurately represented in the next progeny generation when more parents are used for obtaining the mapping population. For this reason, using only one reference family for QTL mapping tends to be less efficient. As information about linkage is accumulated on a within family basis, however, too few sibs per family also gives low power. The optimum structure is dependent on the design used to generate the mapping population.

In practice, due to the great biological flexibility of many fish species, it is possible to utilise a range of other types of mating systems that make use of full sib groups, eg factorial designs. It has been demonstrated previously, however, that different mating designs have little effect on power, because the non-central parameters were essentially the same for the different matings designs considered (Muranty, 1996). The power was more dependent on the number of informative parents that provide informative contrasts than on the number of full sibs families in the population when a constant population size is considered. Therefore, the FS design given here can be considered to represent many alternative designs.

Another design using chromosomal set manipulations in fish is related to the use of meiotic gynogenetic individuals. Essentially, this kind of parthenogenetic reproduction is like selfing, but instead of utilising two different gametes from the same individual to produce progeny, the second polar body is retained. This means that there can be different recombination events in the two chromosomal segments inherited from the same dam, compared with one for the DH design. The expected inbreeding coefficient in this case is equal to 0.5, and it is expected that the variance of the QTL and therefore power increases. The power of QTL detection utilis-

ing selfing has been considered in the context of tree breeding by Kumar *et al* (1999), among others.

So far, the present analysis assumed only additive gene action. Dominance effects can potentially be obtained using factorial designs, whereas for the DH design only additive effects (ie, only homozygote differences) can be estimated. The presence of dominance gene action may increase the power of QTL detection in the FS design, since the square of the marker contrast could include any interaction effects between different alleles segregating (Lynch and Walsh, 1998). More research is needed to investigate the correlation between the performance of outbred and DH individuals coming from the same dam in the presence of dominance.

There are some other practical constraints when applying the DH designs. It has been recognised that inbred individuals often show more environmental variance than non-inbred individuals (Falconer and Mackay, 1996), revealed experimentally by comparing the variance of a trait in inbreds and hybrids. For example in *Drosophila melanogaster*, the variance of wing length increased by about 90% in the inbreds (Robertson and Reeve, 1952). This increase in residual variance may lead to a diminishing of power to detect QTL, since the random background variation is increased. Another study has shown that there is an increase in the within family variance for some quantitative traits compared with the outbred full sib control (Hussain *et al*, 1995). They used only a single family in the experimental design, so it is not straightforward to disentangle increases due to genetic and/or environmental effects. Using the predictions presented here, however, an increase in the environmental variance has a relatively small effect on power. Assuming that the environmental variance is increased by about 50% in the double haploid offspring, the power obtained using the DH design was found still greater than that obtained using the FS design and even when the environmental variance is doubled, the DH design still outperform the FS design (data not shown).

In practice, another possible complication when using populations of homozygous individuals is the presence of segregation distortion due to viability selection (Charlesworth and Charlesworth, 1999). If deleterious mutations are closely linked to marker loci, segregation ratios will be different from those expected under Mendelian inheritance. In QTL mapping studies, this phenomenon has been shown to affect the power of detection as well as producing biased estimates of recombination parameters (Schafer-Pregl *et al*, 1996). Indeed, DH populations from outbred parents may give scope for making inferences about genes associated with fitness and viability traits. Certainly, further research is needed in this area to quantify empirically the magnitude of segregation distortion and how theoretically this may decrease the power when using this type of double haploid populations.

Acknowledgements

VAM gratefully acknowledges financial support from the 'Ministerio de Planificación y Cooperación (MIDEPLAN)' from the Chilean Government and SAK is grateful for support from the Royal Society. We thank Ian White for statistical advice. We also appreciate the constructive criticism of one of the referees.

References

- Bagley MJ, Medrano JF, Gall GAE (1997). Polymorphic molecular markers from anonymous nuclear DNA for genetic analysis of populations. *Mol Ecol* **6**: 309–320.
- Charlesworth D, Charlesworth B (1999). The genetic basis of inbreeding depression. *Genet Res* **74**: 329–340.
- Cheverud JM, Routman EJ, Durante FAM, Van Swinderen, B, Cothran K, Perel C (1996). Quantitative trait loci for murine growth. *Genetics* **142**: 1305–1319.
- Corley-Smith GE, Lim CHJ, Brandhorst BP (1996). Production of androgenetic zebrafish (*Danio rerio*). *Genetics* **142**: 1265–1276.
- Falconer DS, MacKay TFC (1996) *Introduction to Quantitative Genetics*. Fourth Edition. Addison Wesley Longman Limited: Essex, England.
- Geldermann H (1975). Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theor Appl Genet* **46**: 319–330.
- Gomez-Raya L, Sehested E (1999). Power for mapping quantitative trait loci in crosses between outbred lines in pigs. *Genet Sel Evol* **31**: 351–374.
- Haley CS, Knott SA, Elsen JM (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**: 1195–1207.
- Hussain MG, McAndrew BJ, Penman DJ (1995). Phenotypic variation in meiotic and mitotic gynogenetic diploids of Nile tilapia, *Oreochromis niloticus* (L). *Aqua Res* **26**: 205–212.
- Jansen R (1993). Interval mapping of multiple quantitative trait loci. *Genetics* **135**: 205–211.
- Jensen J (1989). Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker loci. *Theor Appl Genet* **78**: 613–618.
- Knapp SJ, Bridges WC, Birkes D (1990). Mapping quantitative trait loci using molecular marker linkage maps. *Theor Appl Genet*, **79**: 583–592.
- Knott SA, Haley CS (1992). Maximum likelihood mapping of quantitative trait loci using full sib families. *Genetics* **132**: 1211–1222.
- Kumar S, Burdon RD, Garrick DJ (1999). Marker-QTL linkage detection in self families of outbred populations. *Silv Genet* **48**: 227–234.
- Luo ZW, Kearsley MJ (1991). Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. II. Application to backcross and double haploid populations. *Heredity* **77**: 117–124.
- Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc: MA, USA.
- Muranty H (1996). Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* **76**: 156–165.
- Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinovitch HD, Lincoln SE *et al* (1991). Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations and environments. *Genetics* **127**: 181–197.
- Robertson FW, Reeve ECR (1952). Heterozygosity, environmental variation and heterosis. *Nature* **170**: 296.
- Robison BD, Wheeler PA, Sundin K, Thorgaard G (2000). Composite interval mapping reveals a QTL of major effect on embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). (abstract). *Proceedings of the VIII Plant and Animal Genome Analysis*, San Diego, California.
- Schafer-Pregl R, Salamini S, Gebhardt C (1996). Models for mapping quantitative trait loci (QTL) in progeny of non-inbred parents and their behaviour in presence of distorted segregation ratios. *Genet Res* **67**: 43–54.
- Streisinger G, Walker CH, Dower N, Knauber D, Singer F (1981). Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* **291**: 293–296.
- Thorgaard GH (1992). Application of genetic technologies to rainbow trout. *Aquaculture* **100**: 85–97.
- Van der Beek S, Van Arendonk JAM, Groen AF (1995). Power of two and three-generation QTL mapping experiments in an

- outbred population containing full-sib or half-sib families. *Theor Appl Genet* **91**: 1115–1124.
- Weller JI, Kashi Y, Soller M (1990). Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci* **73**: 2525–2537.
- Young WP, Wheeler PA, Fields RD, Thorgaard G (1996). DNA Fingerprinting confirms isogenicity of androgenetically derived rainbow trout lines. *J Hered* **87**: 77–81.
- Zeng ZB (1994). Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.