Original article

Power for mapping quantitative trait loci in crosses between outbred lines in pigs

Luis Gomez-Raya*, Erling Sehested

Department of Animal Science, Agricultural University of Norway,
Box 5025, N-1432 Ås, Norway

Abstract – Power for mapping quantitative trait loci using crosses between segregating populations was studied in pigs. Crossing generates gametic disequilibrium and increases heterozygosity. The condition for a heterozygosity among F_1 individuals to be greater than in either line at crossing is that allele frequency should be lower than $1/2$ in one line and higher than $1/2$ in the other line. Maximum expected power and expected risk were used to compare hierarchical backcross (each boar mated to several sows; contrast within boars), hierarchical intercross (each boar mated to several sows; contrast within both boars and sows) and traditional intercross (each boar mated to one sow; contrast within both boars and sows). The use of hierarchical designs (backcross and intercross) increased power for traits with low or intermediate heritabilities. For small QTL effects and low heritabilities the hierarchical backcross design gave the highest expected power but also the highest risk. There is not a general design which allocates resources in an optimum fashion across situations (heritabilities, QTL effect, heterozygosity). A compromise between designs with high power and low risk is suggested. A hierarchical backcross design of 400 piglets and heterozygosity 0.68 requires between four (maximum expected power) and eight boars (minimum risk) to detect a QTL of 0.5 phenotypic standard deviations. Selection of extreme individuals in parental lines increased power up to 21 %. Commercial crosses are proposed as an alternative to experiments for QTL mapping. © Inra/Elsevier, Paris

gene mapping / quantitative trait locus / statistical power / hierarchical design / pig

Résumé – **Puissance de détection des loci à effets quantitatifs dans les croisements entre lignées non consanguines chez le porc.** On a étudié chez le porc la puissance de détection des loci à effets quantitatifs à partir de croisements entre populations en ségrégation. Le niveau d'hétérozygotie chez les individus F_1 est supérieur à celui de

* Correspondence and reprints
E-mail: luis.gomez.raya@akvaforsk.nlh.no

l'une ou l'autre des lignées en croisement si les fréquences alléliques sont supérieures à 1/2 dans l'une des lignées et inférieures à 1/2 dans l'autre lignée. La puissance maximale espérée et le risque espéré ont été considérés pour comparer les croisements en retour hiérarchiques (chaque verrat est accouplé à plusieurs truies : contrastes intra-verrat), l'intercroisement hiérarchique (chaque verrat est accouplé à une truie ; contraste intra-verrat et intra-truie). L'utilisation de ces deux types de schémas hiérarchiques a augmenté la puissance de détection pour les caractères à héritabilités faibles ou intermédiaires. Pour les QTLs à petit effet et les héritabilités faibles, le schéma hiérarchique avec croisement en retour a donné la puissance espérée maximale et aussi le risque le plus élevé. Ce n'est pas le même schéma expérimental qui alloue les ressources d'une façon optimale quand on considère des situations différentes pour l'héritabilité, les effets de QTL et le niveau d'hétérozygotie. Un schéma hiérarchique avec croisement en retour de 400 porcelets et une hétérozygotie de 0,68 requiert entre quatre verrats (puissance espérée maximum) et huit verrats (risque minimal) pour détecter un QTL avec un effet de 0,5 écart-type phénotypique. La sélection d'individus extrêmes dans les lignées parentales a augmenté la puissance jusqu'à 21 %. Les croisements entre lignées commerciales sont proposés comme alternatives aux expérimentations de détection de QTL. © Inra/Elsevier, Paris

détection de gènes / locus à effet quantitatif / puissance statistique / schéma hiérarchique / porc

1. INTRODUCTION

Crosses between inbred lines and analyses within outbred lines have been proposed for quantitative trait loci (QTL) detection. In the first approach, inbred lines are assumed to be fully homozygous for alternative alleles at both marker and QTL. Consequently, F_1 individuals are fully heterozygous and gametic disequilibrium is maximum. High heterozygosity allows high segregation and, therefore, high power. With maximum gametic disequilibrium, a specific allele at the marker is associated with a specific allele at the QTL in all families. Consequently, increased power is achieved because a lower number of contrasts is needed using across family analyses based on a larger number of observations per contrast.

In the second approach, analyses within outbred lines, both marker and QTL are segregating. Power in crosses between inbred lines is much higher than in analyses within outbred lines. A combination of both approaches has assumed that the lines are segregating at the marker but fixed for alternative alleles at the QTL [1, 2]. Another alternative, not explored yet, is to consider parental lines with different allele frequencies at the QTL because of different selection history.

Conventional types of crosses between inbred lines are intercross and backcross. Power of an intercross versus a backcross design is increased because the number of segregating meioses in an intercross is twice as many as in a backcross. Experiments conducted in pigs have restrictions in family structure to a maximum of approximately ten piglets per litter. A small number of full-sibs may result in low power [9]. A hierarchical structure, such as a few boars mated to several sows each, allows larger subgroups of progeny inheriting alternative alleles from each boar, which may increase power.

It has been proposed to use selective genotyping to increase power for QTL detection [10, 11]. Selective genotyping is efficient when the cost of

growing progeny is less than the cost of genotyping. Another use of selection, not considered yet, would be to use extreme individuals for the quantitative trait as parents to produce the F_1. This approach increases the frequency of heterozygotes in the F_1 and, therefore, statistical power.

The objectives of this paper are: 1) to investigate the degree of disequilibrium generated when crossing two populations segregating at two loci; 2) to investigate under which conditions the heterozygosity (and consequently power) in the cross of two lines segregating at a QTL is higher than within either line; 3) to investigate power using hierarchical designs; 4) to investigate optimum allocation of resources for a given experiment size; and 5) to investigate the effect of selection of extreme individuals in the parental lines on heterozygosity and power.

2. THEORY

2.1. Underlying genetic model and assumptions

The purpose of the experimental crossing is the mapping of QTL by the use of genetic markers such as microsatellites. It was assumed that recombination events did not occur between marker and QTL and that all offspring were informative for the marker. These simplifying assumptions were made to reduce the number of parameters to be shown in the results. The impact of these assumptions in practical scenarios is addressed in the discussion.

An underlying mixed inheritance additive model was assumed with an additive biallelic locus segregating at frequencies of the favourable allele p_B and p_C for lines B and C, respectively. The QTL was assumed to be in Hardy-Weinberg equilibrium and in gametic phase equilibrium with the polygene in each parental line. The variance attributable to the QTL in units of the residual phenotypic variance in line i was: $v_Q^2(i) = 2p_i(1 - p_i)\alpha^2/(\sigma_A^2 + \sigma_E^2)$, where p_i is the frequency of the favourable allele in line i, α is the QTL effect, and $(\sigma_A^2 + \sigma_E^2)$ is the residual phenotypic variance with σ_A^2 and σ_E^2 being the additive variance attributable to the polygene component and the environmental variance, respectively. The heritability comprising QTL and polygene variation in parental line i was:

$$h^2(i) = (2p_i(1 - p_i)\alpha^2 + \sigma_A^2)/(2p_i(1 - p_i)\alpha^2 + \sigma_A^2 + \sigma_E^2)$$
$$= (h_r^2 + v_Q^2(i))/(v_Q^2(i) + 1)$$

where h_r^2 is the residual heritability attributable to the polygene component, and has value $\sigma_A^2/(\sigma_A^2 + \sigma_E^2)$. Parameters h_r^2 and v_Q^2 are usually not known and only estimates of $h^2(i)$ are available for some traits. Different heritabilities in the two parental populations were not assumed because of the magnitude of the sampling variance of the estimates of heritability and because parental populations may have been raised under different environmental conditions. For simplicity, power computation in this paper was carried out for a given QTL effect and constant residual heritabilities among F_1 individuals. Therefore, results are valid for a variety of situations with respect to the parental populations segregating at a QTL, not only at a different frequency, but also

with a different heritability (comprising QTL and polygene). The highest possible heritability (QTL and polygene) in the parental populations (denoted as $h^2(\max)$) leading to a given heterozygosity at the QTL among F_1 individuals was computed as an indication of the values of heritability in practical situations. The impact of this assumption on power is addressed in the Discussion. Three levels of heterozygosity in the F_1 were considered: 1, 0.68 and 0.32. A heterozygosity of 1 occurs when the QTL is fixed for alternative alleles. A heterozygosity of 0.68 may occur for a range of situations in parental populations (e.g. $p_B = 0.8$; $p_C = 0.2$ and $p_B = 0.68$; $p_C = 0$). A heterozygosity of 0.32 represents the situation where the two lines segregate at the same frequency (e.g. $p_B = 0.8$; $p_C = 0.8$) or when one allele is absent from one population and segregating at a frequency lower than 0.5 in the other population (e.g. $p_B = 0.32$; $p_C = 0.0$). The advantages of crossing experiments aimed at QTL mapping are: 1) linkage disequilibrium between QTL and marker alleles in the F_1; and 2) increased heterozygosity among F_1 individuals.

2.2. Gametic disequilibrium in the cross of two outbred lines

If the two populations are fixed for alternative alleles at both marker and QTL, then gametic disequilibrium is maximum in the F_1. Therefore, power is increased with respect to within-family analysis. If the two parental populations are segregating at the same frequency, then the resulting F_1 population is in linkage equilibrium and no benefits are expected from crossing. If the genes are segregating at different frequencies in the parental populations then some degree of disequilibrium will be generated. For a more general description of this problem, gametic disequilibrium will be computed at two loci, M and N, that can be either two markers or one marker and one QTL. Gametic disequilibrium between markers can be estimated if estimates of allele frequencies in the parental lines at crossing are available. Let $\mathbf{f}_{bM}$ and $\mathbf{f}_{bN}$ be vectors of allele frequencies at locus M (with k alleles) and at locus N (with g alleles) in parental population B. Similarly, $\mathbf{f}_{cM}$ and $\mathbf{f}_{cN}$ are the corresponding vectors of allele frequencies at loci M and N in parental line C. Assuming that the two loci are in linkage equilibrium within the line then the matrix with gametic frequencies at each pair of alleles for line B is:

$$\mathbf{F}_{bMN} = \mathbf{f}_{bM}\,\mathbf{f}_{bN}' = \begin{bmatrix} \mathbf{f}_{bM1} \\ \mathbf{f}_{bM2} \\ \mathbf{f}_{bM3} \\ \cdot \\ \cdot \\ \mathbf{f}_{bMk} \end{bmatrix} [\mathbf{f}_{bN1}\ \mathbf{f}_{bN2}\ \mathbf{f}_{bN3}\ \cdot\ \cdot\ \mathbf{f}_{bNg}]$$

$$= \begin{bmatrix} \mathbf{f}_{bMN11} & \mathbf{f}_{bMN12} & \mathbf{f}_{bMN13} & \cdot & \cdot & \mathbf{f}_{bMN1g} \\ \mathbf{f}_{bMN21} & \mathbf{f}_{bMN22} & \mathbf{f}_{bMN23} & \cdot & \cdot & \mathbf{f}_{bMN2g} \\ \mathbf{f}_{bMN31} & \mathbf{f}_{bMN32} & \mathbf{f}_{bMN33} & \cdot & \cdot & \mathbf{f}_{bMN3g} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{f}_{bMNk1} & \mathbf{f}_{bMNk2} & \mathbf{f}_{bMNk3} & \cdot & \cdot & \mathbf{f}_{bMNkg} \end{bmatrix}$$

The matrix of gametic frequencies in line C is $\mathbf{F}_{cMN} = \mathbf{f}_{cM}\,\mathbf{f}_{cN}'$. The matrix with gametic frequencies for each pair of alleles at loci M and N in the F_1 of

the cross between lines B and C is given by: $\mathbf{F}_{bcMN} = 1/2(\mathbf{F}_{bMN} + \mathbf{F}_{cMN})$. The allele frequencies for loci M and N in the F_1 cross are $\mathbf{f}_{bcM} = \mathbf{F}_{bcMN}\mathbf{1}$ and $\mathbf{f}_{bcN} = \mathbf{1}'\ \mathbf{F}_{bcMN}$, respectively. In these equations $\mathbf{1}$ represents the unit vector. The gametic disequilibrium matrix is then obtained simply by $\mathbf{D}_{MN} = \mathbf{F}_{bcMN} - (\mathbf{f}_{bcM}\ \mathbf{f}'_{bcN})$. Elements of matrix $\mathbf{D}_{MN}$ are the disequilibrium values for all possible combinations of alleles at loci M and N. Define the disequilibrium parameter between two loci M and N, Ω_{MN}, as the sum of absolute values of all elements of matrix $\mathbf{D}_{MN}$ computed by $\Omega_{MN} = \mathbf{1}'\ \text{abs}(\mathbf{D}_{MN})\mathbf{1}$, where 'abs' denotes the absolute value at each element of the matrix between brackets. Therefore, Ω_{MN} measures the general degree of association between alleles at loci M and N. The value of Ω_{MN} ranges between 0 and 1. For example, Ω_{MN} is 1 when lines at crossing are fixed for alternative alleles and disequilibrium is maximum and Ω_{MN} is 0 when the lines at crossing are segregating at the same frequency for each allele and disequilibrium is null.

2.3. Heterozygosity in the cross of two outbred lines

Consider two outbred lines B and C segregating at a biallelic QTL with alleles Q and q. Assuming random mating, the frequency of heterozygous individuals at the QTL among the F_1 individuals is:

$$he_{BC} = p_B + p_C - 2p_B\,p_C \tag{1}$$

It is shown in Appendix 1 that the necessary condition for increased heterozygosity with respect to either line at crossing is that one line is segregating at a frequency of the favourable allele greater than $1/2$ and the other line at an allele frequency lower than $1/2$.

2.4. QTL mapping designs for the cross between two outbred lines

Power computation was carried out assuming that allele frequencies at the marker were not very different in the two parental populations, and consequently, gametic disequilibrium was small and ignored. However, the increased heterozygosity in the F_1 boars can be used to study segregation, within families, of a QTL associated to a marker by recording performance and inheritance of alternative marker alleles in the next generation (backcross or intercross depending on the mating of F_1 boars to one of the parental lines or F_1 sows, respectively). Therefore, the approach that can be followed is the same as in QTL mapping within outbred lines but with an increased heterozygosity among F_1 individuals. Three alternative designs are considered in this section: hierarchical backcross design, traditional intercross design and hierarchical intercross design. Contrasts in hierarchical designs can use large subgroups of progeny of a mixture of half- and full-sibs within boar.

2.4.1. Hierarchical backcross design

In a hierarchical backcross design, b boars from the F_1 are mated to s sows each to produce p piglets per litter. Sows can be from any parental line. A statistical model to analyse the data could be:

$$y_{ijkl} = bo_i + so_{ij} + m_{ijk} + e_{ijkl} \tag{2}$$

where y_{ijkl} is the lth observation of phenotype on a piglet with marker allele k inherited from boar i when mated to sow j, bo_i is the fixed effect of boar i, so_{ij} is the fixed effect of sow j mated to boar i, m_{ijk} is the fixed effect of marker allele k inherited from boar i, and e_{ijkl} is the residual random error. Model (2) is a three level hierarchical design in which marker alleles are nested within sows, which in turn are nested within boar. For simplicity, sows and boars were assumed fixed. It is not strictly correct because the model does not account for relationships between animals.

Power for model (2) was computed following the χ^2 approach of Gelder-mann [5] and Weller et al. [14] comparing the square of the difference between the two progeny subgroups inheriting alternative alleles from their boar (SDP) for each boar family to the expected squared difference under the null hypothesis. Therefore, the alternative hypothesis is a QTL linked to the marker and the null hypothesis is the absence of a QTL linked to the marker. For a more detailed description of the method and discussion of the assumptions, see Weller et al. [14]. Briefly, the assumptions are complete linkage between a fully informative marker and a QTL, and the analyses are carried out within families. For simplicity, it is assumed that there is no common environmental variance and litter size is fixed at 10.

The statistical tests to reject or to accept the null hypothesis are based on the distribution of the SDP. The summed SDP values divided by the squared standard error of the contrast (SE^2) follow a central χ^2 distribution with b degrees of freedom under the null hypothesis for a large sample size or when phenotypic variance is known. For the alternative hypothesis $\Sigma(\text{SDP}/\text{SE}^2) \sim \chi^2(nc, b)$, where nc is the non-centrality parameter of a non-central χ^2 distribution with b degrees of freedom. The value of the non-centrality parameter is $nc = b\, he_{BC}\, \theta^2/\text{SE}^2$, where $\theta = \alpha + \delta(1 - 2p_B)$ and $\theta = \alpha + \delta(1 - 2p_C)$, for backcrosses with lines B and C, respectively (Appendix 2), α is half the difference between the two alternative homozygotes, δ is the dominance deviation, and $\text{SE}^2 = 4\left[(1 + (1/4)\text{sh}_\text{r}^2 - 1/2\text{h}_\text{r}^2)/(sp)\right]$ [equation (A3) in Appendix 3].

Following Weller et al. [14], the power to detect a segregating QTL was computed as $1 - \beta = 1 - p[\chi^2(nc, b) < T]$, where β is the probability of committing a type 2 error, $[p(\chi^2(nc, b) < T]$ is the probability of a χ^2 value under the alternative hypothesis (non-central χ^2 with parameter nc and b degrees of freedom) less than T, with T being the value of the central χ^2 (b) for a given significance level of committing a type 1 error.

2.4.2. Traditional intercross design

In the intercross design, inbred lines B and C are crossed to produce the F_1 boars and sows that are intercrossed among themselves. In the traditional intercross design b boars are mated to one sow each. A linear model allowing testing for the marker–QTL effects is:

$$y_{ijk} = f_i + m_{ij} + e_{ijk} \tag{3}$$

where y_{ijk} is the kth observation on piglet k with marker genotype j inherited in family i, f_i is the fixed effect of family i, m_{ij} is the fixed effect of marker–QTL genotype j inherited in family i ($j = 1$ to 3; homozygous for either allele

and heterozygous) and e_{ijk} is the residual error. Model (3) would be the one to use in the analysis of real data. However, to take advantage of the χ^2 approach of Geldermann [5] and Weller et al. [14] for power calculation, the following linear model can be used:

$$y_{ijkl} = f_i + mb_{ij} + ms_{ik} + e_{ijkl} \tag{4}$$

where y_{ijkl} is the lth phenotypic observation of piglet inheriting marker allele k from the sow and allele j from the boar, mb_{ij} is the fixed effect of marker allele j inherited from the boar, ms_{ik} is the fixed effect of marker allele k inherited from the sow, and e_{ijkl} is the residual error.

Model (4) is linearly equivalent to model (3) under the assumption of a fully additive underlying genetic model and no sexual imprinting. The distribution under the alternative hypothesis of $\Sigma(\text{SDP}/\text{SE}^2)$ is approximately $\chi^2(nc, b)$ and $\chi^2(nc, s)$ for boars and sows, respectively. Assuming that the hypothesis being tested is the same either within boars or within sows, the distribution of the sum of $\Sigma(\text{SDP}/\text{SE}^2)$ for both boars and sows follows a $\chi^2(2nc, b+s)$ under the alternative hypothesis. The sum of variables having non-central χ^2 distributions jointly independent also follows a non-central χ^2 distribution with degrees of freedom equal to the sum of the degrees of freedom of the former non-central χ^2 and non-centrality parameter equal to the sum of the non-centrality parameters of the former variables having non-central χ^2. The non-centrality parameter of the χ^2 has a value $2nc = (b + s) \, he_{BC} \, \theta^2/\text{SE}^2$, where $\theta = \alpha + \delta(1 - p_B - p_C)$, and $\text{SE}^2 = 4[(1 - (1/2)h_r^2)/p]$ [equation (A4) in Appendix 3]. Similarly to a hierarchical backcross design, power to detect a segregating QTL was computed as $1 - \beta = 1 - p[\chi^2(2nc, 2b) < T]$, where β is the probability of committing a type 2 error, $[p(2(2nc, 2b) < T]$ is the probability of a χ^2 value under the alternative hypothesis (non-central χ^2 with parameter $2nc$ and $2b$ degrees of freedom) less than T, with T being the value of the central χ^2 $(2b)$ for a given significance level of committing a type 1 error.

2.4.3. Hierarchical intercross design

In a hierarchical intercross design, b boars from the F_1 are mated to s sows each from the F_1 to produce p piglets per litter. Power calculation can be carried out using the linear model:

$$y_{ijklm} = bo_i + so_{ij} + mb_{ik} + ms_{ijl} + e_{ijklm} \tag{5}$$

where y_{ijklm} is the mth observation of phenotype on a piglet with marker allele k inherited from boar i mated to sow j, bo_i is the fixed effect of boar i, so_{ij} is the fixed effect of sow j mated to boar i, mb_{ik} is the fixed effect of marker allele k inherited from boar i, ms_{ijl} is the fixed effect of marker allele l inherited from sow ij and e_{ijklm} is the residual error. Power calculation using model (5) requires different non-centrality parameters within boars and sows. The non-centrality parameter in boars is: $nc_b = b \, he_{BC}\theta^2/\text{SE}^2$, where $\theta = \alpha + \delta(1 - p_B - p_C)$ (Appendix 2), with $\text{SE}^2 = 4[(1 + (1/4)ph_r^2 - 1/2h_r^2)/(sp)]$ for $s > 1$ (Appendix 3). Values of SE^2 for $s = 1$ are as for traditional intercross

design. The non-centrality parameter in sows is $nc_s = b\,s\,he_{BC}\theta^2/\text{SE}^2$, where $\theta = \alpha+\delta(1-p_B-p_C)$ (Appendix 2), and $\text{SE}^2 = 4[(1-(1/2)h_r^2)/p]$ (Appendix 3). Assuming independence, the distribution of the sum of $\Sigma(\text{SDP}/\text{SE}^2)$ for both boars and sows follows a $\chi^2(nc_b+nc_s,\ b(1+s))$ under the alternative hypothesis. Under the null hypothesis, the distribution of the sum of $\Sigma(\text{SDP}/\text{SE}^2)$ for both boars and sows follows a central $\chi^2(b(1+s))$. Power to detect a segregating QTL was computed as $1-\beta = 1 - p[\chi^2(nc_b + nc_s,\ b(1+s)) < T]$, where β is the probability of committing a type 2 error, $[p(\chi^2(nc_b+nc_s,\ b(1+s)) < T]$ is the probability of a χ^2 value under the alternative hypothesis (non-central χ^2 with parameter $nc_b + nc_s$ and $b(1+s)$ degrees of freedom) less than T, with T being the value of the central $\chi^2\,(b(1+s))$ for a given significance level of committing a type 1 error.

2.5. Design of experiments using crosses between outbred lines

Most experimental costs are in raising, genotyping and recording of a given number of slaughter pigs in the F_2 or backcross. Therefore, the parameters to be chosen by the researcher are the number of boars, the number of sows and the number of piglets per sow. In practice, it is convenient for the handling of the experiment to fix the number of piglets per sow. It is also economical since the use of few piglets per sow would increase the cost in raising a large number of sows. It will be assumed from now on that the number of piglets per litter in the experiment is fixed at ten and the question is how to make optimal use of different numbers of boars and sows in the F_1 for a given experiment size (total number of slaughter pigs).

Parameters of interest are the expected power (EP) and its standard deviation (SD):

$$\text{EP} = \sum_{i=0}^{b} pr(i)\,P_i$$

$$\text{SD} = \left[\sum_{i=0}^{b} pr(i)\,P_i^2 - \text{EP}^2\right]^{1/2}$$

where $pr(i)$ is the probability according to the binomial distribution of having i heterozygous boars [assuming a frequency of heterozygous boars given by equation (1)], and P_i is the power with i boars computed according to the previous sections. Expected power is utilized for each design to account for random sampling of the boars. SD can be used as a measurement of the variation in power. Comparison of alternative crossing designs can be done by comparing expected power and its standard deviation at a fixed experiment size. The approach used in this paper for optimum allocation of resources is based on the repeated computation of expected power for all possible combinations of the number of boars and sows at a given experiment size assuming constant litter size of ten piglets. The power of the mating structure having the highest expected power will be called maximum expected power (MEP). The functions CINV and PROBCHI of SAS [12] were used to compute central and non-central χ^2 probabilities, respectively.

Another parameter of interest in planning experiments is expected risk (ER) which we define as the probability of having power lower than 0.5 due to sampling among F_1 individuals in the segregating population:

$$\text{ER} = \sum_{i=1}^{b} pr(i)\, f(x) \text{ where } f(x) = 1 \text{ if } P_i < 0.5, \text{ and } f(x) = 0, \text{ otherwise}$$

2.6. Selection in parental lines to increase power for QTL mapping

Selection of high ranking individuals in one parental line and of low ranking individuals in another parental line can be used to increase statistical power for the experiment when the lines are not fixed for alternative alleles. Assume a biallelic QTL segregating in the two parental lines. Consider a hierarchical backcross design in which high and low ranking individuals selected in the parental lines are randomly mated to produce F_1, which are again randomly mated to one of the unselected parental lines. Two effects would occur in this scheme: 1) an increase in the frequency of heterozygous individuals in the F_1 which would increase the power to detect QTL; and 2) an increase in genetic variance of the backcross offspring due to linkage disequilibrium [3], which would decrease power to detect QTL.

2.6.1. Frequency of heterozygotes among offspring of high and low ranking parents

Computation of the frequency of heterozygous F_1 individuals was carried out by using integrals simultaneously of the three normal distributions corresponding to the three genotypes at the QTL. The selected proportion (p_s) for the high line is related to the proportion selected among individuals with genotypes QQ, Qq and qq $(p_{QQ}, p_{Qq}$ and $p_{qq})$ by the equation:

$$p_s = p_B^2\, P_{QQ} + 2p_B(1 - p_B)p_{Qq} + (1 - p_B)^2\, p_{qq}$$

where
$$p_{QQ} = \int_t^\infty f(x|G_i = QQ), \ p_{Qq} = \int_t^\infty f(x|G_i = Qq)$$

and
$$p_{qq} = \int_t^\infty f(x|G_i = qq)$$

with $f(x|G_i)$ being the normal density given genotype $(G_i = QQ, Qq$ or $qq)$. Values p_{QQ}, p_{Qq}, and p_{qq} can be found for a unique truncation point, t, analytically. In the examples considered in this paper Simpson's rule [13] was used for increasing values of the abscissa until a value of t was found satisfying the above equation for a given p_s. Computation of expected genotype frequencies among high ranking individuals after selection in the high line was carried out by

$$p_{QQ,H}^* = p_B^2 p_{QQ}/p_s, \ p_{Qq,H}^* = 2p_B(1 - p_B)\, p_{Qq}/p_s$$

and
$$p_{qq,H}^* = (1 - p_B)^2\, p_{qq}/p_s$$

The same procedure was followed to obtain expected genotype frequencies among low ranking individuals in the low line $(p^*_{QQ,L}, p^*_{Qq,L}, p^*_{qq,L})$ using the same proportion selected, p_s, as for the high line. Computing the expected frequency of heterozygous F_1 offspring resulting from the cross of selected parents from each line was carried out with equation (3) of Gomez-Raya and Gibson [6]. It assumes random mating and utilizes the frequencies among extreme individuals from each of the two parental lines.

2.6.2. Increased variance among the offspring of F_1 individuals

The effect of selection of high and low ranking individuals on genetic variance and on the estimation of heritability is discussed by Bulmer [4] and Gomez-Raya et al. [7]. Briefly, if selection of high and low ranking individuals is carried out in the same population then genetic variance increases in the selected group of parents $\sigma^2_{AS} = (1 + k_s r^2)\sigma^2_A$, where σ^2_A is the genetic variance for the polygenic component in the unselected population, $k_s = [z\,\phi(z)]/p_s$, z is the absolute value of standard normal deviates at cutoff points, $\phi(z)$ is the ordinate and r is accuracy of evaluation. The above equation can also be used when crossing high ranking individuals from one parental line with low ranking individuals from the other parental line as long as the parental populations have the same genetic variance and the same polygenic heritability and allowance is made for the different selection intensities for each genotype at the QTL and for each line. For simplicity, an approximation to account for different selection pressures across genotypes was made by weighting the values of z and $\phi(z)$ according to the proportion selected from each genotype and line in the computation of k_s by

$$k_s = \left[\sum_{i=1}^{3} p^*_{i,H} z^H_i \phi(z^H_i) + p^*_{i,L} z^L_i \phi(z^L_i)\right] \bigg/ 2p_s$$

where z^m_i and $\phi(z^m_i)$ are the absolute values of standard normal deviate and ordinate at truncation point for genotype i and line m (i had values 1, 2 and 3 for genotypes QQ, Qq and qq, respectively).

An exact computation would be feasible by using all possible groups resulting in the F_1 offspring and by including the variation of the means corresponding to the different groups. This variation is small, particularly for QTL of small effect, and the approach used accounts for the increased variance among offspring from selected parents.

Following Bulmer [3], the genetic variance in the F_1 generation is $\sigma^2_{A(F_1)} = (1 + (1/2)k_s r^2)\sigma^2_A$. The disequilibrium is halved in the next generation since parents in the F_1 are chosen randomly. The genotypic variance among backcross offspring from randomly mating F_1 individuals to one of the parental lines (unselected) is:

$$\sigma^2_{A(BC)} = (1/2)\sigma^2_A + (1/2)\sigma^2_{A(F1)} = (1 + (1/4)k_s r^2)\sigma^2_A$$

The heritability among backcross offspring is also increased by selection of high and low ranking individuals in the parental lines:

$$h^2_{BC} = [(1 + (1/4)k_s r^2)\sigma^2_A]/[(1 + (1/4)k_s r^2)\sigma^2_A + \sigma^2_E]$$
$$= [(1 + (1/4)k_s r^2)h^2_r]/(1 + (1/4)k_s r^2 h^2_r)$$

Power calculation was carried out as described in the corresponding section but using the increased frequency of heterozygotes and heritability as computed above. The estimates of the gene effect are biased if selection of parents is used. Correction for selection bias in the estimates of the gene effect can be achieved after accounting for the increased variance among backcross offspring by

$$\widehat{\alpha} = \widehat{\alpha}^* \sqrt{\frac{\sigma_A^2 + \sigma_E^2}{\sigma_{A(BC)}^2 + \sigma_E^2}} = \widehat{\alpha}^* \frac{1}{\sqrt{1 + (1/4)k_s r^2 h_r^2}}$$

where $\widehat{\alpha}^*$ is the estimate of the gene effect from the experiment using selection, and $\widehat{\alpha}$ is the estimate of the gene effect after correction for selection bias. Both $\widehat{\alpha}^*$ and $\widehat{\alpha}$ are in phenotypic standard deviations units.

3. RESULTS

Table I shows the gametic disequilibrium parameter for alternative number of alleles segregating in the parental populations. The disequilibrium is high when alleles at each of the two loci are fixed in one population but they are rare in the other population. If the alleles at the two loci are segregating at a similar frequency in the two populations then disequilibrium is small. In this situation, the analysis could be performed within families allowing for different alleles at the marker to be associated with the same allele at the QTL in different families.

Maximum expected power and its standard deviation for three different crossing designs in a variety of situations (QTL effect, residual heritability, experiment size) for heterozygosity 1, 0.68 and 0.32, are given in *tables II, III* and *IV*, respectively. The heritability including QTL and polygenic variation (h^2(max)) of the parental population having the highest possible value is given in these tables. It can be observed that residual heritability has a small effect on power for the range of heritabilities in the parental populations. Therefore, power figures can be considered as a good approximation when accurate estimates of heritability in the parental population are not available.

On the other hand, a reduction in the frequency of heterozygotes among F_1 individuals diminished the power for any situation considered. For example, for residual heritability 0.2 and QTL effect 0.5 in an experiment with 200 piglets in a hierarchical backcross design, the maximum expected power with a significance level of 0.05 fell from 0.85 to 0.31 for heterozygosity 1 and 0.32, respectively. Power using hierarchical designs (backcross and intercross) increases in all cases with the exception of traits with high heritability when compared to traditional intercross designs (*tables II–IV*). Maximum expected power using hierarchical backcross designs is larger than using hierarchical intercross designs when the QTL effect is small. The opposite occurs for large QTL effects.

The use of a larger experiment size increases power in all cases. For example, the average maximum power across QTL size and residual heritability (*table II*) for heterozygosity 1 in hierarchical backcross designs is 0.60, 0.79 and 0.91 for experiment sizes 200, 400 and 800, respectively. Similarly, for hierarchical

Table I. Disequilibrium parameter, Ω_{MN}, in the F_1 cross of two populations, B and C, segregating at alternative allele frequencies for loci M and N. The elements of vectors $\mathbf{f}_{cM}$ and $\mathbf{f}_{cN}$ are allele frequencies for loci M and N in population C, respectively. Similarly, $\mathbf{f}_{bM}$ and $\mathbf{f}_{bN}$ are vectors of allele frequencies for loci M and N in population B.

			Pop C		
		alleles	2	2	4
		$\mathbf{f}'_{cM}$	$[0\ 1\ 0\ldots0]$	$[1/2\ 1/2\ 0\ldots0]$	$[1/4\ 1/4\ 1/4\ 1/4\ 0\ldots0]$
		$\mathbf{f}'_{cN}$	$[1\ 0\ 0\ldots0]$	$[1/2\ 1/2\ 0\ldots0]$	$[1/4\ 1/4\ 1/4\ 1/4\ 0\ldots0]$
	alleles	$\mathbf{f}_{bM}\quad\mathbf{f}_{bN}$			
	2	$\begin{bmatrix}1\\0\\0\\.\\.\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\.\\.\\0\end{bmatrix}$	1		
Pop B	2	$\begin{bmatrix}1/2\\1/2\\0\\.\\.\\0\end{bmatrix}\begin{bmatrix}1/2\\1/2\\0\\.\\.\\0\end{bmatrix}$	0.25	0	
	4	$\begin{bmatrix}1/4\\1/4\\1/4\\1/4\\1/4\\0\\.\\.\\0\end{bmatrix}\begin{bmatrix}1/4\\1/4\\1/4\\1/4\\0\\.\\.\\0\end{bmatrix}$	0.56	0.25	0
	10	$\begin{bmatrix}1/10\\1/10\\1/10\\.\\.\\1/10\end{bmatrix}\begin{bmatrix}1/10\\1/10\\1/10\\.\\.\\1/10\end{bmatrix}$	0.81	0.64	0.36

intercross designs, average maximum power is 0.58, 0.71 and 0.81 for experiment sizes of 200, 400 and 800, respectively.

Variation in power due to the sampling in the F_1 should also be considered. Hierarchical backcross designs show a much larger variation than either intercross designs (*tables III* and *IV*). An example of how power, standard deviation of power and expected risk change for alternative mating designs is given in *table V*. In this example, heterozygosity is 0.68, residual heritability of the trait is 0.2 and QTL effect is 0.5 phenotypic standard deviations with experiment size of 400 piglets. The resulting expected power utilizing a hierarchical backcross design is maximum when using two boars and 20 sows (0.79). The

Table II. Maximum expected power for detection of an additive QTL using hierarchical backcross designs (HBC), traditional intercross design (TID) and hierarchical intercross designs (HIC) at a significance level of 0.05 for different residual heritability (h_r^2), and QTL effect in phenotypic standard deviations (α). Heterozygosity in the F_1 is 1. The experiment sizes are 200, 400 and 800 piglets and the number of piglets per litter is ten. The highest possible value of heritability (QTL and polygene) in either parental population is h^2(max).

h_r^2	α	h^2 (max)	200 piglets			400 piglets			800 piglets		
			HBC	TID	HIC	HBC	TID	HIC	HBC	TID	HIC
	0.2	0.1	0.25	0.12	0.15	0.45	0.17	0.21	0.73	0.24	0.31
	0.3	0.1	0.49	0.27	0.33	0.78	0.41	0.51	0.97	0.62	0.75
0.1	0.4	0.1	0.73	0.51	0.60	0.95	0.75	0.84	1.00	0.94	0.98
	0.5	0.1	0.90	0.77	0.85	1.00	0.95	0.98	1.00	1.00	1.00
	0.6	0.1	0.97	0.94	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.2	0.22	0.13	0.14	0.39	0.17	0.20	0.67	0.25	0.30
	0.3	0.2	0.43	0.28	0.32	0.72	0.43	0.49	0.95	0.66	0.73
0.2	0.4	0.2	0.67	0.54	0.58	0.92	0.78	0.83	1.00	0.96	0.97
	0.5	0.2	0.85	0.80	0.83	0.99	0.97	0.98	1.00	1.00	1.00
	0.6	0.2	0.95	0.95	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.5	0.17	0.15	0.15	0.29	0.21	0.21	0.52	0.31	0.31
	0.3	0.5	0.32	0.35	0.35	0.56	0.53	0.53	0.85	0.78	0.78
0.5	0.4	0.5	0.52	0.64	0.64	0.81	0.88	0.88	0.98	0.99	0.99
	0.5	0.5	0.71	0.89	0.89	0.94	0.99	0.99	1.00	1.00	1.00
	0.6	0.5	0.85	0.98	0.98	0.99	1.00	1.00	1.00	1.00	1.00

corresponding standard deviation of power and expected risk are 0.27 and 10 %, respectively. A compromise between expected power and risk can be taken. For example, the use of five boars has a low expected risk of 4 % with an expected power of 0.74. It is more obvious with a hierarchical intercross design where expected power using between one and ten boars is very similar (0.81 to 0.83). However, the expected risk is only low when using four or more boars.

Hierarchical and traditional intercross designs gave identical results at heritability 0.5 (*tables II–V*). This is because traditional intercross is an extreme case of hierarchical intercross (one sow per boar) which is the best allocation of resources at high heritabilities.

Table VI shows the number of boars, expected power and risk for designs having maximum expected power and minimum risk in hierarchical intercross and backcross designs. The residual heritabilities are 0.10, 0.20 and 0.50 and the QTL size ranges from 0.3 to 0.6 phenotypic standard deviations in an experiment of 400 piglets with F_1 heterozygosity of 0.68. It can be observed that the numbers of boars for maximum expected power and minimum risk are different in many instances but the values are not very far apart. It should also be noticed that the optimal allocation of resources in the experiment varies with the QTL effect. A small number of boars is required for small QTL effects which in turn would also have high risk.

Table III. Maximum expected power for detection of an additive QTL using hierarchical backcross designs (HBC), traditional intercross design (TID) and hierarchical intercross designs (HIC) at a significance level of 0.05 for different residual heritability (h_r^2), QTL effect in phenotypic standard deviations (α). Heterozygosity in the F_1 is 0.68. The experiment sizes are 200, 400 and 800 piglets and the number of piglets per litter is ten. Heritability comprising QTL and polygenic variation (h^2(max)) corresponds to the parental line having the highest possible heritability for heterozygosity of 0.68 in the F_1 (e.g. line B when $p_B = 0.68$; $p_C = 0$). The values between brackets are standard deviations of expected power.

h_r^2	α	h^2(max)	200 piglets			400 piglets			800 piglets		
			HBC	TID	HIC	HBC	TID	HIC	HBC	TID	HIC
0.1	0.2	0.12	0.19 (0.09)	0.10 (0.01)	0.11 (0.02)	0.32 (0.19)	0.12 (0.01)	0.15 (0.04)	0.51 (0.32)	0.16 (0.01)	0.20 (0.06)
	0.3	0.13	0.35 (0.21)	0.18 (0.03)	0.22 (0.07)	0.55 (0.34)	0.26 (0.03)	0.33 (0.11)	0.74 (0.26)	0.40 (0.04)	0.51 (0.17)
	0.4	0.16	0.51 (0.32)	0.33 (0.06)	0.41 (0.13)	0.70 (0.26)	0.51 (0.06)	0.60 (0.14)	0.89 (0.16)	0.74 (0.05)	0.83 (0.08)
	0.5	0.19	0.63 (0.40)	0.54 (0.09)	0.62 (0.19)	0.82 (0.27)	0.78 (0.07)	0.84 (0.09)	0.96 (0.10)	0.95 (0.02)	0.98 (0.02)
	0.6	0.22	0.74 (0.26)	0.75 (0.09)	0.80 (0.13)	0.91 (0.15)	0.94 (0.04)	0.96 (0.04)	0.99 (0.03)	1.00 (0.00)	1.00 (0.00)
0.2	0.2	0.21	0.17 (0.08)	0.10 (0.01)	0.11 (0.02)	0.28 (0.16)	0.12 (0.01)	0.14 (0.03)	0.47 (0.29)	0.17 (0.01)	0.20 (0.05)
	0.3	0.23	0.31 (0.18)	0.19 (0.03)	0.21 (0.06)	0.50 (0.31)	0.27 (0.03)	0.32 (0.09)	0.69 (0.26)	0.42 (0.04)	0.49 (0.15)
	0.4	0.25	0.47 (0.29)	0.35 (0.06)	0.39 (0.12)	0.65 (0.25)	0.54 (0.06)	0.59 (0.17)	0.85 (0.17)	0.78 (0.05)	0.82 (0.08)
	0.5	0.28	0.59 (0.37)	0.57 (0.09)	0.60 (0.17)	0.79 (0.27)	0.81 (0.06)	0.83 (0.08)	0.94 (0.11)	0.97 (0.02)	0.97 (0.02)
	0.6	0.31	0.69 (0.26)	0.78 (0.09)	0.79 (0.12)	0.88 (0.16)	0.95 (0.03)	0.96 (0.04)	0.98 (0.05)	1.00 (0.00)	1.00 (0.00)
0.5	0.2	0.51	0.13 (0.06)	0.11 (0.01)	0.11 (0.01)	0.22 (0.11)	0.14 (0.01)	0.14 (0.01)	0.37 (0.22)	0.20 (0.02)	0.20 (0.02)
	0.3	0.52	0.24 (0.13)	0.23 (0.04)	0.23 (0.04)	0.40 (0.24)	0.34 (0.04)	0.34 (0.04)	0.59 (0.37)	0.53 (0.05)	0.53 (0.05)
	0.4	0.53	0.37 (0.22)	0.43 (0.07)	0.43 (0.07)	0.57 (0.35)	0.65 (0.07)	0.65 (0.07)	0.76 (0.27)	0.88 (0.04)	0.88 (0.04)
	0.5	0.55	0.50 (0.31)	0.68 (0.09)	0.68 (0.09)	0.68 (0.25)	0.90 (0.05)	0.90 (0.05)	0.87 (0.16)	0.99 (0.01)	0.99 (0.01)
	0.6	0.57	0.59 (0.37)	0.87 (0.07)	0.87 (0.07)	0.79 (0.27)	0.99 (0.01)	0.99 (0.01)	0.94 (0.11)	1.00 (0.00)	1.00 (0.00)

Table IV. Maximum expected power for detection of an additive QTL using hierarchical backcross designs (HBC), traditional intercross design (TID) and hierarchical intercross designs (HIC) at a significance level of 0.05 for different residual heritability (h_r^2), QTL effect in phenotypic standard deviations (α) and heterozygosity of 0.32. Heritability comprising QTL and polygenic variation ($h^2(\text{max})$) corresponds to the parental line having the highest possible heritability for heterozygosity of 0.32 in the F_1 (e.g. line B when $p_B = 0.32$; $p_C = 0$). The experiment sizes are 200, 400 and 800 piglets and the number of piglets per litter is ten. The values between brackets are standard deviations of expected power.

h_r^2	α	h^2 (max)	200 piglets			400 piglets			800 piglets		
			HBC	TID	HIC	HBC	TID	HIC	HBC	TID	HIC
0.1	0.2	0.12	0.11 (0.09)	0.07 (0.01)	0.08 (0.02)	0.18 (0.19)	0.08 (0.01)	0.09 (0.03)	0.27 (0.32)	0.09 (0.01)	0.11 (0.05)
	0.3	0.13	0.19 (0.21)	0.10 (0.02)	0.12 (0.06)	0.28 (0.34)	0.13 (0.02)	0.16 (0.09)	0.42 (0.35)	0.17 (0.03)	0.23 (0.16)
	0.4	0.16	0.27 (0.32)	0.15 (0.04)	0.20 (0.12)	0.39 (0.33)	0.21 (0.05)	0.28 (0.19)	0.56 (0.31)	0.32 (0.06)	0.41 (0.20)
	0.5	0.19	0.34 (0.28)	0.24 (0.08)	0.30 (0.20)	0.49 (0.32)	0.35 (0.09)	0.43 (0.21)	0.69 (0.31)	0.54 (0.10)	0.63 (0.18)
	0.6	0.22	0.42 (0.35)	0.35 (0.13)	0.42 (0.20)	0.59 (0.32)	0.53 (0.13)	0.61 (0.19)	0.79 (0.25)	0.77 (0.10)	0.83 (0.12)
0.2	0.2	0.21	0.11 (0.08)	0.07 (0.01)	0.08 (0.02)	0.16 (0.16)	0.08 (0.01)	0.09 (0.03)	0.25 (0.29)	0.09 (0.01)	0.11 (0.04)
	0.3	0.23	0.17 (0.18)	0.10 (0.02)	0.12 (0.05)	0.26 (0.31)	0.13 (0.02)	0.15 (0.08)	0.38 (0.32)	0.18 (0.03)	0.22 (0.13)
	0.4	0.25	0.25 (0.29)	0.16 (0.05)	0.19 (0.10)	0.36 (0.30)	0.23 (0.06)	0.27 (0.17)	0.51 (0.34)	0.34 (0.07)	0.40 (0.18)
	0.5	0.28	0.31 (0.37)	0.25 (0.09)	0.29 (0.18)	0.46 (0.38)	0.38 (0.10)	0.42 (0.19)	0.64 (0.34)	0.58 (0.10)	0.62 (0.16)
	0.6	0.31	0.38 (0.32)	0.38 (0.13)	0.40 (0.24)	0.55 (0.35)	0.56 (0.13)	0.59 (0.18)	0.74 (0.31)	0.80 (0.10)	0.82 (0.12)
0.5	0.2	0.51	0.09 (0.06)	0.07 (0.01)	0.07 (0.01)	0.13 (0.11)	0.09 (0.01)	0.09 (0.01)	0.20 (0.22)	0.11 (0.01)	0.11 (0.01)
	0.3	0.52	0.14 (0.13)	0.12 (0.03)	0.12 (0.03)	0.21 (0.24)	0.15 (0.03)	0.15 (0.03)	0.31 (0.37)	0.22 (0.04)	0.22 (0.04)
	0.4	0.53	0.20 (0.22)	0.19 (0.06)	0.19 (0.06)	0.29 (0.35)	0.28 (0.07)	0.28 (0.07)	0.43 (0.36)	0.43 (0.08)	0.43 (0.08)
	0.5	0.55	0.26 (0.31)	0.31 (0.11)	0.31 (0.11)	0.38 (0.32)	0.46 (0.12)	0.46 (0.12)	0.54 (0.35)	0.69 (0.11)	0.69 (0.11)
	0.6	0.57	0.31 (0.37)	0.46 (0.16)	0.46 (0.16)	0.46 (0.38)	0.67 (0.14)	0.67 (0.14)	0.64 (0.34)	0.89 (0.08)	0.89 (0.08)

Table V. Expected power (EP), standard deviation (SD), and expected risk (ER) in hierarchical backcross designs (HBD) and hierarchical intercross designs (HIC) for different numbers of boars (b) and sows per boar (s). The significance level is 0.05. The size of the experiment is 400. The heterozygosity in the F_1 is 0.68. The residual heritability of the trait is 0.2 and the QTL effect is 0.5 phenotypic standard deviations.

		HBC			HIC		
b	s	EP	SD	ER*	EP	SD	ER*
1	40	0.69	0.44	32.0	0.81	0.17	5.7
2	20	0.79	0.27	10.2	0.83	0.11	1.8
4	10	0.76	0.18	10.0	0.83	0.08	0.2
5	8	0.74	0.16	3.9	0.83	0.08	0.1
8	5	0.67	0.13	7.5	0.82	0.06	0.0
10	4	0.64	0.12	18.7	0.81	0.06	0.0
20	2	0.49	0.08	46.9	0.77	0.05	0.0
40	1	0.35	0.04	100.0	0.81	0.06	0.0

* Values have been multiplied by 100.

Table VI. Expected power (EP) and expected risk (ER) for population structures (b boars in the F_1) having maximum expected power (MEP) or minimum risk (MR) in hierarchical backcross designs (HBD) and hierarchical intercross designs (HID). The significance level is 0.05. The size of the experiment is 400 piglets. The residual heritabilities (h_r^2) are 0.1, 0.2 and 0.5. The heterozygosity among F_1 individuals is 0.68. Heritability comprising QTL and polygenic variation (h^2(max)) corresponds to the parental line having the highest possible heritability for heterozygosity of 0.68 in the F_1 ($p_B = 0.68$; $p_C = 0$). The QTL effect in phenotypic standard deviations (α) ranges from 0.3 to 0.6.

		HBC						HIC					
		b	EP	ER*	b	EP	ER*	b	EP	ER*	b	EP	ER*
$h_r^2 = 0.1$	h^2(max)												
$\alpha = 0.3$	0.13	1	0.55	32.0	1	0.55	32.0	1	0.33	100.0	1	0.33	100.0
$\alpha = 0.4$	0.16	2	0.70	10.2	4	0.64	10.0	2	0.60	18.0	8	0.57	14.6
$\alpha = 0.5$	0.19	2	0.82	10.2	5	0.80	3.9	4	0.84	0.4	20	0.79	0.0
$\alpha = 0.6$	0.22	4	0.91	1.0	20	0.77	0.9	5	0.96	0.0	8	0.96	0.0
$h_r^2 = 0.2$													
$\alpha = 0.3$	0.23	1	0.50	32.0	1	0.50	32.0	1	0.32	100.0	1	0.32	100.0
$\alpha = 0.4$	0.25	2	0.65	10.2	2	0.65	10.2	1	0.59	32.0	5	0.57	19.2
$\alpha = 0.5$	0.28	2	0.79	10.2	5	0.74	3.9	4	0.83	0.2	20	0.77	0.0
$\alpha = 0.6$	0.31	4	0.88	1.0	4	0.88	1.0	5	0.96	0.0	5	0.96	0.0
$h_r^2 = 0.5$													
$\alpha = 0.3$	0.52	1	0.40	32.0	1	0.40	32.0	40	0.34	100.0	40	0.34	100.0
$\alpha = 0.4$	0.53	1	0.57	32.0	1	0.57	32.0	40	0.65	2.9	40	0.65	2.9
$\alpha = 0.5$	0.55	2	0.68	10.2	2	0.68	10.2	40	0.90	0.0	40	0.90	0.0
$\alpha = 0.6$	0.57	2	0.79	10.2	5	0.74	3.9	40	0.99	0.0	40	0.99	0.0

* Values have been multiplied by 100.

Heterozygosity and power using phenotypic selection of extreme individuals in a hierarchical backcross design with 400 piglets, residual heritability 0.20 and varying percentage selected are depicted in *figures 1* and *2*, respectively. Parental lines were segregating at frequencies of 0.8 (selected to increase) and 0.2 (selected to decrease). A percentage selected of 100 is given for comparison with the unselected case. Percentage selected of 0.5 % was the lowest attempted. Heterozygosity increases with increasing selection pressure in each line and with the size of the QTL. Following the same pattern, power increases with decreasing percentage selected being up to 21 % more (*figure 2*). The increase is small if the power in the unselected situation is close to 0 or 1.

Percentage Selected in Each Line
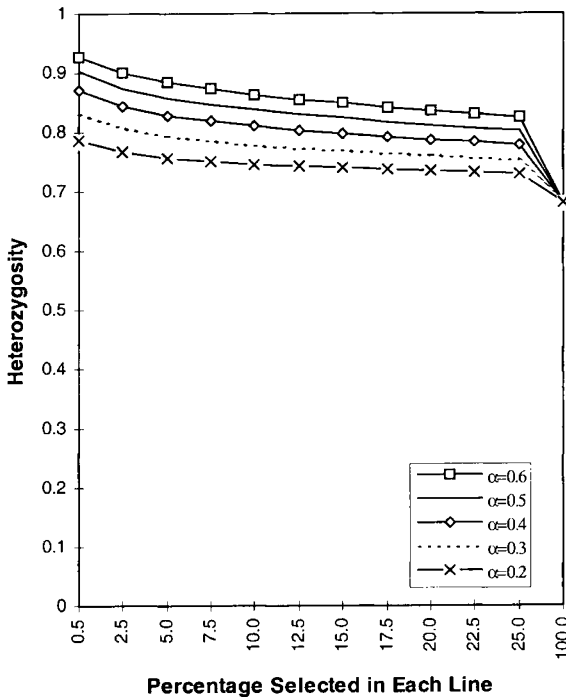
Figure 1. Values of heterozygosity for varying percentage selected in each line when crossing lines with allele frequencies 0.8 and 0.2, respectively. Phenotypic selection of extreme individuals was for increasing the trait in the line with frequency 0.8 and for decreasing the trait in the line with frequency 0.2. The QTL effect (α) was 0.2, 0.3, 0.4, 0.5 and 0.6 phenotypic standard deviation. The residual heritability was 0.2. Heritability comprising QTL and polygenic variation in parental lines was 0.21, 0.22, 0.24, 0.26 and 0.28 for QTL effects 0.2, 0.3, 0.4, 0.5 and 0.6, respectively.

4. DISCUSSION

Most previous research in QTL mapping has assumed that lines at crossing are fixed for alternative alleles at the QTL [1, 2, 8]. The availability of divergent inbred lines for traits of economical interest is a limiting factor in these studies.
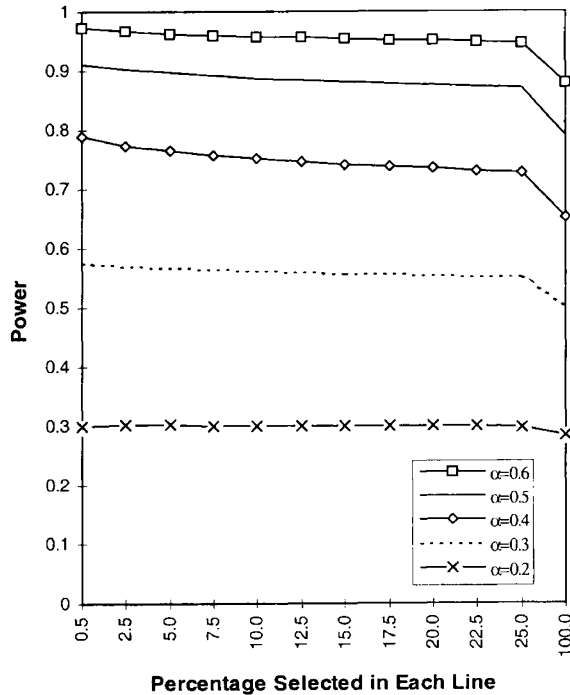
Figure 2. Maximum expected power, at a significance level of 0.05, for varying percentage selected in each line in a hierarchical backcross design when crossing lines with allele frequencies 0.8 and 0.2, respectively. Phenotypic selection of extreme individuals was for increasing the trait in the line with frequency 0.8 and for decreasing the trait in the line with frequency 0.2. The QTL effect (α) was 0.2, 0.3, 0.4, 0.5 and 0.6 phenotypic standard deviation. The residual heritability was 0.2. Heritability comprising QTL and polygenic variation in parental lines was 0.21, 0.22, 0.24, 0.26 and 0.28 for QTL effects 0.2, 0.3, 0.4, 0.5 and 0.6, respectively.

Domesticated breeds of pigs have been undergoing artificial selection to increase growth rate and litter size, for example. Only for research purposes, selection criterion has been low growth rate or reduced litter size in experimental populations. Therefore, it seems more reasonable to assume that the lines at crossing differ in allele frequency at the QTL rather than that they are fixed for alternative alleles.

The benefits of crossing lines are to increase gametic disequilibrium between marker and QTL alleles and to increase heterozygosity among F_1 individuals. Gametic disequilibrium generated at crossing is high when the allele frequency at the two loci (either two markers or one marker and one QTL) is very different in the two parental lines. Gametic disequilibrium between two linked markers could be estimated in chromosomal fragments by the use of the disequilibrium parameter proposed in this paper. It would require that markers are not tightly linked so parental lines could be assumed to be in linkage equilibrium. If the gametic disequilibrium is low (e.g. $\Omega < 0.2$) then the analysis could be carried

out within families as considered in this paper with only a small loss in power. If the gametic disequilibrium is high (e.g. $\Omega > 0.8$) then a maximum likelihood approach could be developed allowing the same marker allele to be associated with the same QTL allele only in some families. Consequently, power figures as given in this study represent a lower bound of the achievable power in those cases. More work is needed to assess the gain in power in situations where disequilibrium parameter has intermediate values.

Power for QTL detection in experiments involving crosses between two outbred lines is higher than using within-line experiments when the frequency of the favourable allele is higher than $1/2$ in one parental line and lower than $1/2$ in another parental line. The increased power can be attributed to the higher frequency of heterozygous F_1 individuals than in either parental breed. The larger the difference in allele frequency between the two parental populations, the larger is the increase in power.

It was assumed in the computation of power that residual heritability was constant for varying QTL effects and for a given heterozygosity in the F_1. This represents a variety of situations in which parental populations can be segregating for a QTL at a different frequency and having a different heritability. The approach taken in this paper was to compute also the heritability (comprising QTL and polygene variation) of the parental line with the highest possible value for a given heterozygosity in the F_1. At low residual heritability, the contribution of large QTL to the heritability is high. For example, for residual heritability 0.1 and QTL effect of 0.6 phenotypic standard deviations, the heritability is 0.22 when the allele frequency in the parental population is 0.68 (table III). In spite of the large contribution of the QTL variation to the heritability, changes in power for different heritability values are small. Discrepancies in power comparing residual heritability 0.1 and 0.2 for the same QTL size range between 0.00 and 0.05 (table III). Therefore, power figures, as given in this paper, are well approximated when precise estimates of heritabilities from parental population are not available. The approximation is particularly good when the QTL has a small effect.

On the other hand, it was assumed when computing power that recombination between QTL and marker did not occur. In most instances, recombination would occur. If the recombination fraction between the marker and the QTL is c then the number of offspring should be increased in proportion to $1/(1-2c)^2$ to obtain the same power as with complete linkage [14]. Power could be increased by using interval mapping instead of the use of a single marker [10]. Most of the relevant information in interval mapping comes from the non-recombinant individuals. The number of offspring required to obtain power as given in this paper should be increased in a proportion $1/(1-c')$, where c' is the recombination fraction between the two markers. That is, around 25 % more offspring are needed to achieve power as given in this paper for interval mapping between two markers with a recombination fraction of 0.20. A larger experiment size will also be needed to achieve the same power figures as given in this paper when the marker is not fully informative. The use of interval mapping can mitigate the lack of informative offspring by utilizing information corresponding to nearby informative flanking markers.

The models used in this paper ignore the possibility of a common full-sib family component such as may occur for some traits of economic interest in

pigs. Power for mapping QTL affecting those traits would be reduced [9]. The approach used in this study to compute power could be modified to account for this problem by incorporating a term corresponding to the common family component in the standard error of the contrasts.

It has also been assumed that the heritabilities in parental populations as well as within the offspring of F_1 remain unchanged. The frequency of heterozygotes at other loci affecting the trait could increase genetic variance among F_1 individuals and, therefore, heritability. One generation of random mating is enough to restore Hardy-Weinberg equilibrium as would occur after intercrossing F_1 individuals. However, changes in the genetic variance among offspring resulting from the hierarchical backcross design may occur. If there are other QTL affecting the trait showing dominance or epistasis, then changes in the phenotypic variance and heritability could also occur. Therefore, departures from the power as computed in this paper would be expected for traits that show heterosis such as reproductive traits.

For a given power, a backcross requires about twice as many progeny as an intercross design because the number of segregating meioses is doubled in the intercross [10]. Experimental design for mapping QTL in pigs has the limitation that the number of piglets per litter has a maximum around ten. This limits the size of the subgroups of progeny where the contrasts are carried out. Hierarchical backcross and intercross designs can, however, be used in pigs and allow contrasts within boars with larger subgroups of progeny. Power of intercross versus backcross depends on both heritability and QTL effect when using hierarchical designs. At high heritability levels, the hierarchical intercross design is more powerful than the hierarchical backcross design. The opposite occurs at low heritability values.

An interesting result was that power for small QTL effects was higher in a hierarchical backcross than in a hierarchical intercross. The only difference between the two designs is that the hierarchical backcross ignores meioses from sows. In a hierarchical intercross, contrasts between alternative alleles inherited from sows are small with a large variation given the small QTL effect and the small progeny group (ten piglets). However, the total number of degrees of freedom may be high because of the large number of sows. On the contrary, in a hierarchical backcross each of the few boars has large subgroups of progeny which means that contrasts have low variation and hypothesis testing is carried out with a low number of degrees of freedom. As a consequence of the above, power may be reduced in the intercross.

There is not a general experimental design which can allocate resources in an optimum way if the experiment size is fixed. Power in experiments with a low number of boars is higher for small size of QTL and high heritability (*table VI*). The opposite occurs for QTL with large size and low heritability. Risk measured as the probability of power lower than 0.5 due to sampling of F_1 boars is high with a low number of boars. If the frequency of heterozygous F_1 individuals is low then the risk increases and the power decreases (results not shown). A compromise between expected power and risk should be made. A possibility for experiments searching for QTL affecting traits with known heritability is first to decide the potential QTL size and F_1 heterozygosity detectable for a given experiment size (*tables II–IV*). Second is to choose a design for that QTL size with maximum expected power but conditional on an

expected risk of less than a value of, say, 0.05. For example, for traits with low heritability and using hierarchical backcross designs with an experiment size of 400 piglets, the detectable QTL sizes are 0.3, 0.3 and 0.5 for a heterozygosity of 1, 0.68 and 0.32, respectively (*tables II–IV*). Between four (maximum expected power) and eight boars (minimum risk) could be chosen for a heterozygosity of 0.68 and a QTL effect of 0.5 phenotypic standard deviations (*table VI*). If the heterozygosity is lower then the QTL will be very difficult to detect even using a design with optimal allocation of resources.

Dam and sire lines currently utilized for commercial production of pigs (e.g. Large White, Landrace and Duroc, in Norway) can be considered as parental lines which are crossed to produce F_1 individuals with high frequency of heterozygotes. Genotyping both boars (F_1) and piglets (backcross to Landrace) and recording of traits in the slaughter pigs can be used to map QTL by contrasting inheritance of alternative alleles with large subgroups of progeny. Utilizing both sire and dam lines would also be feasible in a similar fashion to a hierarchical intercross design. However, the cross between F_1 boars and sows is not currently used by the industry. The swine industry in other countries often uses four-way crosses. For example, line A is crossed to line B to produce $F_1(AB)$, line C is crossed to line D to produce $F_1(CD)$. The cross $F_1(AB) \times F_1(CD)$ would also show increased heterozygosity and power with respect to experiments within line if the allele frequency is higher than 1/2 among $F_1(AB)$ and lower than 1/2 among $F_1(CD)$ individuals and vice versa. The use of commercial pigs for QTL mapping has two advantages: 1) saving in the cost of running the experiment since the carcasses can be sold at current prices in the market; and 2) findings can immediately be used in marker-assisted selection in the commercial stocks.

Selective genotyping involving growing a large population but genotyping only those individuals whose phenotypes deviate far from the mean has been proposed to increase power for QTL detection [10, 11]. In this paper an alternative use of selection to increase power is proposed. Selection of extreme individuals in the parental lines segregating at a QTL can be used to increase heterozygosity in the F_1 and consequently, power of the experiment. It has been shown that selection is more useful with experimental designs having intermediate values of power without selection. The increase in power is up to 21 %. A higher power would be feasible if high selection intensities are practised, for example, by screening large populations for extreme individuals. The main restriction with either use of selection (extreme individuals in the parental lines or extreme progeny) is that in many instances, QTL mapping experiments are carried out for several traits, which makes impractical the use of selection.

ACKNOWLEDGEMENTS

REFERENCES

[1] Andersson L., Haley C.S., Ellegren H., Knott S.A., Johanson M., Andersson K., Andersson-Eklund L., Edfors-Lilja I., Fredholm M., Hanson I., Håkanson J., Lundström K., Genetic mapping of quantitative trait loci for growth and fatness in pigs, Science 263 (1994) 1771–1774.

[2] Beckmann J.S., Soller M., Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations, Theor. Appl. Genet. 76 (1988) 228–236.

[3] Bulmer M.G., The effect of selection on genetic variability, Am. Nat. 105 (1971) 201–211.

[4] Bulmer M.G., The effect of selection on genetic variability: a simulation study, Genet. Res. 28 (1976) 101–117.

[5] Geldermann H., Investigations on inheritance of quantitative characters in animals by gene markers. 1. Methods, Theor. Appl. Genet. 46 (1975) 319–357.

[6] Gomez-Raya L., Gibson J.P., Within-family selection at a specific locus in dairy cattle, Genome 36 (1993) 433–439.

[7] Gomez-Raya L., Schaeffer L.R., Burnside E.B., Selection of sires to reduce sampling variance in the estimates of heritability by half-sib correlation, Theor. Appl. Genet. 81 (1991) 624–628.

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

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

[10] Lander E.S., Botstein D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, Genetics 121 (1989) 185–199.

[11] Lebowith R.J., Soller M., Beckmann J.S., Trait based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines, Theor. Appl. Genet. 73 (1987) 556–562.

[12] SAS Language: Reference, Version 6, 1990, First Edition, Cary, NC, SAS Institute Inc., 1042 p.

[13] Smith W.A., Elementary Numerical Analysis, Prentice-Hall, New Jersey, 1986.

[14] Weller J.I., Kashi Y., Soller M., Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle, J. Dairy Sci. 73 (1990) 2525–2537.

APPENDIX 1: Heterozygosity in the cross of two outbred lines

In this section, the conditions for increased heterozygosity in the cross between two segregating populations at a QTL is studied. Consider two outbred lines B and C segregating at a biallelic QTL with a favourable allele frequency p_B and p_C for lines B and C, respectively. Under Hardy-Weinberg equilibrium, the frequency of heterozygous individuals at the QTL in breed B is:

$$he_B = 2p_B(1 - p_B)$$

Similarly, the frequency of heterozygotes among F_1 individuals in the cross is

$$he_{BC} = p_B + p_C - 2p_B\,p_C$$

Rearranging the above equation

$$he_{BC} = 2p_B(1 - p_B) + d(2p_B - 1) \qquad (A1)$$

where $d = p_B - p_C$, the difference in allele frequency between the two breeds.

The amount $d(2p_B - 1)$ is the excess of heterozygotes in the cross with respect to line B. The same reasoning can be used to obtain the frequency of heterozygotes in the cross as a function of d and p_C:

$$he_{BC} = 2p_C(1 - p_C) + d(1 - 2p_C) \qquad (A2)$$

The following can be concluded by inspection of equations A1 and A2.

i) If $p_C > 1/2$ and $p_B > 1/2$ then $d(2p_B - 1)$ is positive and $d(1 - 2p_C)$ is negative.

ii) If $p_C < 1/2$ and $p_B < 1/2$ then $d(2p_B - 1)$ is negative and $d(1 - 2p_C)$ is positive.

For i or ii, the heterozygosity in the offspring of a cross is not higher than in the parental line (B or C) with the highest heterozygosity.

iii) If $p_C > 1/2$ and $p_B < 1/2$ then $d(2p_B - 1)$ is positive and $d(1 - 2p_C)$ is positive.

iv) If $p_C < 1/2$ and $p_B > 1/2$ then $d(2p_B - 1)$ is positive and $d(1 - 2p_C)$ is positive.

For iii or iv, the heterozygosity in the offspring of a cross is higher than the heterozygosity within either line, B or C.

Therefore, crossing between outbred lines segregating at a QTL with a frequency larger than $1/2$ in one line and lower than $1/2$ in the other leads to increased heterozygosity in F_1, which increases statistical power for QTL mapping with respect to analysis within line.

APPENDIX 2: Contrasts in backcross and intercross designs

Outbred lines B and C are segregating at a biallelic QTL (with alleles Q and q) with favourable allele (Q) at frequencies p_B and p_C for lines B and C, respectively. Offspring from a heterozygous (Qq) F_1 boar inheriting allele Q can be QQ (with genetic value α) and Qq (with genetic value δ). Offspring inheriting q can be qQ (with genetic value δ) and qq (with genetic value $-\alpha$). The genotype frequencies of each type of offspring depending on the mating of F_1 individuals with parental line B (backcross BC-B), parental line C (backcross BC-C), or with other F_1 individuals (intercross) are:

Allele inherited from boar	Genotype offspring	Genetic value	Frequency BC-B	BC-C	Intercross
Q	QQ	α	$(1/2)p_B$	$(1/2)p_C$	$(1/2)((1/2)p_B + (1/2)p_C)$
Q	Qq	δ	$(1/2)(1 - p_B)$	$(1/2)(1 - p_C)$	$(1/2)(1 - (1/2)p_B - (1/2)p_C)$
q	qQ	δ	$(1/2)p_B$	$(1/2)p_C$	$(1/2)((1/2)p_B + (1/2)p_C)$
q	qq	$-\alpha$	$(1/2)(1 - p_B)$	$(1/2)(1 - p_C)$	$(1/2)(1 - (1/2)p_B - (1/2)p_C)$

Contrasts within boar for backcrosses with lines B and C are $\theta = \alpha + \delta(1 - 2p_B)$ and $\theta = \alpha + \delta(1 - 2p_C)$, respectively. Contrast for the intercross is $\theta = \alpha + \delta(1 - p_B - p_C)$. Therefore, for fully additive models (as assumed in the paper) $\theta = \alpha$.

APPENDIX 3: Standard error of contrasts

The following derivations are for a phenotypic variance of 1. Consequently, other variances are expressed as a proportion of the phenotypic variance. It is also assumed that there is no residual covariance between sires. Consider first a hierarchical backcross design where y_{ijkl} is the lth observation of phenotype on a piglet with marker allele k inherited from boar i mated to sow j. The total variance is the variance of the means of the subgroups of progeny (mixed full- and half-sibs) being contrasted and has value

$$
\begin{aligned}
V_T &= V((\Sigma_j \Sigma_l y_{ijkl})/sp) \\
&= (sp + (p-1)((1/2)h_r^2)sp + sp^2(s-1)(1/4)h_r^2)/(sp)^2 \\
&= [(1 + (1/4)sh_r^2 - (1/2)h_r^2)/(sp)] + (1/4)h_r^2
\end{aligned}
$$

where h_r^2 is the residual heritability, s is the number of sows per boar and p is the number of piglets per sow. The total variance of the means is also $V_T = V_b + V_w$, where V_b and V_w are the components between and within families, respectively. The standard error of the contrast in a hierarchical cross design is two times the square root of V_w since the contrast is within the progeny groups inheriting alternative alleles from their boar and half of the total number of daughters per boar would receive each allele.

The square of the standard error can be computed by:

$$
\begin{aligned}
\mathrm{SE}^2 &= 2^2 V_w \\
&= 4(V_T - V_b) \\
&= 4[(1 + (1/4)sh_r^2 - (1/2)h_r^2)/(sp)] \tag{A3}
\end{aligned}
$$

since $V_b = (1/4)h_r^2$.

The same approach can be used for the traditional intercross design. In this case the total variance is given by

$$
\begin{aligned}
V_T &= V((\Sigma_l y_{ijk})/p) \\
&= [(1 - (1/2)h_r^2)/p] + (1/2)h_r^2
\end{aligned}
$$

where y_{ijk} is the kth observation on piglet k with marker genotype j inherited in family i. The standard error of the contrast in the traditional intercross design is

$$
\begin{aligned}
\mathrm{SE}^2 &= 2^2 V_w \\
&= 4[(1 - (1/2)h_r^2)/p] \tag{A4}
\end{aligned}
$$

Note that the traditional intercross design is a particular case of a hierarchical intercross design with one sow per boar. Therefore, SE^2 was computed using A4 when $s = 1$ in hierarchical intercross designs.