J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

# Fine mapping quantitative trait loci under selective phenotyping strategies based on linkage and linkage disequilibrium criteria

S. Ansari-Mahyari[1,2,3], P. Berg[1] & M.S. Lund[1]

1 Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, University of Aarhus, Tjele, Denmark
2 Agricultural Biotechnology Research Institute of Iran (ABRII), Karaj, Iran
3 Department of Large Animal Sciences, Faculty of Life Sciences, University of Copenhagen, Copenhagen, Denmark

**Keywords**
linkage disequilibrium; phenotyping criteria; quantitative trait loci.

**Correspondence**
S. Ansari-Mahyari, Department of Genomics, Branch of North region of Iran, Agricultural Biotechnology Research Institute of Iran (ABRII), P.O. Box 31535-1897, Karaj, Iran.
Tel: +98 261 2709652;
Fax: +98 261 2701067;
E-mail: sam@abrii.ac.ir

Received: 3 July 2008;
accepted: 23 March 2009

## Summary

In fine mapping of a large-scale experimental population where collection of phenotypes are very expensive, difficult to record or time-demanding, selective phenotyping could be used to phenotype the most informative individuals. Linkage analyses based sampling criteria (LAC) and linkage disequilibrium-based sampling criteria (LDC) for selecting individuals to phenotype are compared to random phenotyping in a quantitative trait loci (QTL) verification experiment using stochastic simulation. Several strategies based on LAC and LDC for selecting the most informative 30%, 40% or 50% of individuals for phenotyping to extract maximum power and precision in a QTL fine mapping experiment were developed and assessed. Linkage analyses for the mapping was performed for individuals sampled on LAC within families and combined linkage disequilibrium and linkage analyses was performed for individuals sampled across the whole population based on LDC. The results showed that selecting individuals with similar haplotypes to the paternal haplotypes (minimum recombination criterion) using LAC compared to random phenotyping gave at least the same power to detect a QTL but decreased the accuracy of the QTL position. However, in order to estimate unbiased QTL parameters based on LAC in a large half-sib family, prior information on QTL position was required. The LDC improved the accuracy to estimate the QTL position but not significantly compared to random phenotyping with the same sample size. When applying LDC (all phenotyping levels), the estimated QTL effect were closer to the true value in comparison to LAC. The results showed that the LDC were better than the LAC to select individuals for phenotyping and contributed to detection of the QTL.

## Introduction

In detecting and mapping loci affecting quantitative traits (QTL), the association between the phenotype and DNA markers extending across the genome is assessed in a genetically segregating population. Due to recent developments in high-throughput technologies, genotyping costs are generally less limiting to the sample size in QTL mapping experiments. The genotyping facilities have drastically reduced the cost of genotyping (Jobs *et al.* 2003). This is particularly true for single nucleotide

© 2009 Blackwell Verlag GmbH  doi:10.1111/j.1439-0388.2009.00813.x

polymorphism markers (Bell *et al.* 2002). In contrast to these developments for genotyping, there are some situations where recording the phenotypes of traits is difficult or very expensive and therefore, a question of how to find the most informative individuals to phenotype in fine mapping experiments has appeared.

Darvasi & Soller (1992) introduced a genotyping sampling method for identifying individuals for genotyping called selective genotyping (SG). This method selected the most informative individuals to genotype based on phenotype (up to half of the population) and the power of QTL detection was as large as genotyping the whole population (100% genotyping). The SG may be advantageous if expenses of rearing are fairly low and the trait is routinely evaluated. Instead of considering the phenotype to identify informative individuals in genotyping, selective phenotyping strategies (SP) are necessary for situations where costs of phenotyping are larger than costs of genotyping or the trait measurements cannot be easily collected. Medugorac & Soller (2001) and Casu *et al.* (2003) practised SP when a main trait was highly correlated with an indicator trait which was easily recorded and showed that using a correlated trait can be useful in QTL detection wherever the main trait is difficult or expensive to record. These studies indirectly used the SP based on information from the correlated trait.

Several SP studies have used information from linkage analyses between the parents and their offspring (Jannink 2005; Jin *et al.* 2004), where most information is provided by highly recombinant individuals to increase genetic dissimilarity in the sample sets. In fact, these methods proposed a set of linkage criteria which were based on prior knowledge of QTL position and maximum recombinations in the phenotyped individuals. Therefore, if more information about the genetic architecture can be available and the approximate region containing the QTL of interest is assumed to be known, shorter intervals may be considered in a QTL fine mapping experiment with a large population. Therefore, after detecting the QTL based on an initial genome scan (Darvasi 1998), SP can be used to obtain further accuracy in QTL fine mapping studies.

Linkage disequilibrium (LD) information across a population can utilize information on historical recombinations of alleles at different loci in an identified chromosomal segment, which can be used to select and genotype/phenotype informative animals to fine map QTL, for example with half-sib family designs in outbred populations (Hayes *et al.* 2006) or granddaughter designs as in Olsen *et al.* (2004), who utilized linkage disequilibrium and linkage analyses (LDLA) in dairy cattle based on the identity by descent (IBD) method of Meuwissen *et al.* (2002). Statistical methods based on all available information and considering joint linkage and LD analyses have been reported in QTL experiments of open-pollinated populations (Wu & Zeng 2001; Wu *et al.* 2002). Moreover, LDLA may be particularly useful to confirm association of candidate regions that have already been shown to include QTL.

The objective of this study was to develop criteria for ranking animals for phenotyping based on genotype information in a QTL fine mapping experiment for a typical outcrossing livestock population. The criteria were based on information from linkage analyses or LD across a chromosomal segment. The sampling criteria were examined in a set of half-sib families using simulated data and different proportions of phenotyped individuals were studied and compared to a random phenotyping approach.

## Materials and methods

### Outline of the simulation

A stochastic simulation was set up to create LD among six highly polymorphic markers and a QTL over 99 generations of random mating in a historical population without recording the pedigree and with constant population size of 100 males and 100 females in each generation (see Meuwissen & Goddard 2000). The genotypes for all individuals were simulated for marker loci equally spaced in a 10 cM chromosomal segment and also an associated QTL placed in the midpoint between markers 2 and 3. A simple bi-allelic QTL and markers with five alleles were assumed. In the first generation, the alleles were sampled and the frequencies of alleles at each marker were equal to 0.2. In the following 99 generations, marker and QTL genotypes were sampled according to rules of Mendelian inheritance and allowed for recombination within the segment. In generation 100, a mutant QTL allele was sampled at random with the requirement that the frequency of the positive QTL allele ($P_Q$) was between 0.45 and 0.55, otherwise the LD population was re-simulated.

Generation 101 was defined as base population with unknown parents (founders). Marker genotypes were assumed known for parents (generation 101) and marker genotypes plus pedigree and phenotypic

values were known in the progenies (generation 102). It was assumed that recombinants could be identified unambiguously, with known genotype phase. A quantitative trait with moderate inheritance ($h^2 = 0.25$) was modelled with total phenotypic variance of one. Proportion of QTL variance relative to the total phenotypic variance was between 0.062 and 0.063 [$\sigma^2_{QTL} = 2 \times P_Q \times (P_Q - 1) \times \delta^2$], where $\delta$ is the additive allele substitution effect ($\delta = \mu_{QQ} - \mu_{qq}$) and assumed to be 0.354. Polygenic and residual effects were drawn from $N(0, \boldsymbol{A}\sigma^2_a)$ and $N(0, \boldsymbol{I}\sigma^2_e)$, where $A$ is the numerator relationship matrix, which ignored the ancestral relationship beyond the known pedigree, $\boldsymbol{I}$ is an identity matrix, and $\sigma^2_a$ and $\sigma^2_e$ are polygenic and residual variance components, respectively. Polygenic effects of the daughters were equal to means of the paternal and maternal values plus a Mendelian sampling component. Mendelian effects were sampled from $N(0, \frac{1}{2}\sigma^2_a)$. After generating an observation, the QTL effect was added. In this study, 30 unrelated sires were simulated with 80 daughters per sire. Therefore, progenies (daughters) were paternal half-sibs and all dams were assumed unrelated.

### Phenotyping criteria

In order to choose the most informative daughters, two sources of information in a daughter design were considered, based on linkage information within each sire family (linkage analyses based sampling criteria, LAC) and LD across half-sib families (linkage disequilibrium-based sampling criteria, LDC). Midpoints of the five marker brackets ($N_{pos} = 5$) were considered the putative QTL positions. All markers on both sides of a putative QTL position were considered as a haplotype, such that the haplotypes included six markers but with a varying number of markers on the left and right side of the putative QTL position.

*Random phenotyping*

With this method, individuals were randomly phenotyped either within sire families or across the families. Three levels of phenotyping (30, 40 and 50%), either an equal number of randomly chosen daughters within sire families or a random sample across sire families, were considered.

*Linkage analyses based sampling criteria (LAC)*

This approach used linkage information by following the recombination events within each half-sib family. Therefore, marker genotypes in the parents and daughters were considered.

*Maximum recombinants (MaxRec)*: This strategy maximizes recombination across the haplotype and samples the most recombinant offspring given paternal genotypes. For each progeny ($j$) and sire ($k$), define $R_{ijk} = 1$ if the progeny is recombinant in marker interval $i$ in the paternal inherited genotype, and $R_{ijk} = 0$ otherwise.

The number of recombinant marker intervals for progeny $j$ within sire $k$ is:

$$R_{.jk} = \sum_{i=1}^{Npos} R_{i,j,k} \quad (1)$$

The $N$ progeny with most recombinations were selected for phenotyping within sire family groups.

*Maximum-uniform recombinants (MaxUniRec)*: While the previous strategy maximizes the overall recombination among the selected progeny in each sire family, MaxUniRec samples daughters with many recombinations, such that the recombinations across sampled individuals are uniformly distributed over all marker intervals. The rationale for this strategy followed from the assumption that the QTL position was unknown across the segment and therefore, it was desirable to have mapping information evenly distributed. The sampling was performed in two stages. In the first stage, half the candidates were selected based on the maximum recombinations (as in the previous strategy) leading to 12, 16 and 20 of the highest recombinant progenies within each sire group being phenotyped. Therefore in total for 30 sire families, 360 daughters in 30% phenotyping; 480 daughters in 40% phenotyping; and 600 daughters in 50% phenotyping were selected in the first stage based on maximum recombination. All selected daughters were assigned to the set $S$. In the second stage, the following steps were iterated until $N$ progeny were selected (720, 960 and 1200 selected for the levels of 30, 40 and 50%, respectively):

**(i)** The remaining daughters within each sire family were added to $S$, one by one, and recombination events were summed for the five intervals (the intervals of six markers). Then, (SD) of recombinations across the intervals was calculated as:

$$SD = \sqrt{\frac{1}{N_{pos} - 1}\sum_{i=1}^{N_{pos}} (x_i - \bar{x})^2};$$

where, $N_{pos}$ as described above, $x_i$ is the total number of recombinations in interval $i$, and $\bar{x}$ is the

mean number of recombinations across intervals. This approach was continued for all sire families to compute changes in SD when one new daughter was added to *S*.

**(ii)** In the next step, within each sire family, the daughters resulting in the lowest SD were added to *S*. If several progeny had the same SD, one was chosen at random.

*Minimum recombination (MinRec)*: The aim of this criterion was to sample paternal haplotypes and increase power by sampling as many records as possible per haplotype. This strategy minimizes recombination events across the intervals in the paternal haplotype of half-sib progeny. The score was eqn 1, the number of recombinant marker intervals for progeny *j* within sire *k*. Then, progeny with the lowest recombination score within the sire families were selected for phenotyping. The same numbers of progeny were phenotyped in each family. Therefore, this strategy primarily samples non-recombinant offspring for phenotyping.

*Linkage disequilibrium-based sampling criteria (LDC)*

In order to use LD information within an outcross population e.g. half-sib families, several criteria were applied. The main objective in these criteria was to phenotype informative individuals across the families for QTL detection based on historical recombinations. Therefore, depending on the putative QTL location across the segment, the parental haplotypes were clustered (as detailed by Meuwissen & Goddard 2001).

*Maximum frequency (MaxFre)*: This approach aims to maximize frequencies of the clustered haplotypes of the founders (parents) while minimizing recombinations in offspring. Therefore, first all unique haplotypes in the founders were identified. After clustering the haplotypes in founders according to Meuwissen & Goddard (2001), the frequency of clustered haplotypes was assessed. In each particular interval, the cluster group of the paternal and maternal haplotypes could be identified in their offspring based on the marker genotype information. Then over all putative QTL positions, the frequencies of paternal and maternal haplotypes (HapFre) were summed in the offspring as below. Equation 2 was used when a recombination is observed in the marker bracket (the interval assumed to include a QTL), and Equation 3 if not:

$$\mathrm{HapFre}_{\mathrm{individual}} = \sum_{i\mathrm{pos}=1}^{N\mathrm{pos}} \left[\frac{1}{2}(P_1 + P_2) + \frac{1}{2}(M_1 + M_2)\right] \quad (2)$$

$$\mathrm{HapFre}_{\mathrm{individual}} = \sum_{i\mathrm{pos}=1}^{N\mathrm{pos}} [(P + M)] \quad (3)$$

where, $P_1$ and $P_2$ are frequencies of paternal and maternal clustered haplotypes in the sire, and $M_1$ and $M_2$ in the dam, and $P$ is paternal and $M$ is maternal clustered haplotype frequencies of sire and dam, respectively. Ultimately, the offspring with the highest HapFre across sire families in the population were chosen for phenotyping at three levels of phenotyping.

*Equal highly frequent haplotypes (EqHigh)*: This strategy minimizes the variance of the frequency of clustered haplotypes among the phenotyped offspring. Therefore, the objective is to equalize the frequency of the most frequent clusters in each putative QTL position and sample a balanced set of frequent clustered haplotypes. In order to start this strategy, first the optimum number of cluster groups was found by minimizing the standard error of the variance component according to a one way random model (Searle *et al.* 1992). A preliminary study showed that using 50 of the most frequent clustered haplotypes was optimal ($N_{\mathrm{hap}}$ = 50). Before beginning this sampling criterion, two steps were needed. In step one, 50 of the most frequent haplotypes were found in each putative QTL position, based on clustering the haplotypes in parents (founders). Then, the first 10% of the population were phenotyped as the second step. These phenotyped daughters were chosen such that clustered haplotypes of their parents should have maximum similarity with the most frequent haplotype clusters already selected. The next daughters were iteratively chosen for phenotyping based on the following index until the proposed phenotyping level of the population was sampled:

$$\mathrm{Index} = \sum_{i\mathrm{pos}=1}^{N\mathrm{pos}} \left( \sum_{i\mathrm{hap}=1}^{N\mathrm{hap}} \left( N_{i\mathrm{pos},i\mathrm{hap}} - \frac{2 \times N_{\mathrm{animal}}}{50} \right)^2 + N_{\mathrm{unsel}}^2 \right) \quad (4)$$

where, $N_{\mathrm{pos}}$ is the number of putative QTL positions across the segment, $N_{i\mathrm{pos},i\mathrm{hap}}$ is the number of the most frequent haplotype cluster in each position, $N_{\mathrm{animal}}$ is the number of offspring to be phenotyped in each level of phenotyping, $N_{\mathrm{unsel}}$ is the number of haplotype clusters not in the list of most frequent haplotypes. If the haplotype codes of the candidates


© 2009 Blackwell Verlag GmbH 

existed in the codes derived in the first initial step, then $N^2_{\text{unsel}}$ was set to zero. Therefore, $N^2_{\text{unsel}}$ can be different for each candidate, according to their founder haplotypes.

*Maximum similar highly frequent clusters (Max)*: In this strategy, daughters are selected for phenotyping if their parental haplotypes have high similarity with highly frequent haplotypes in the population. First, all five QTL positions are recognized. Then according to the founder genotypes, the most frequent haplotype clusters are identified. If parental haplotypes in the candidates had maximum similarity with these highly frequent haplotype clusters over all the positions, the candidates are chosen to be phenotyped. This process continued until the three levels of phenotyping were achieved. Two sizes of this sampling criterion were used in order to evaluate the size effect of the most frequent haplotype clusters and also robustness of the criterion. In Max10, only 10 of the most frequent clusters in each position were assigned and in Max50, 50 of the most frequent clustered haplotype codes.

### Analyses of the simulated data

In total 100 data sets were simulated and the sampling criteria and three levels of phenotyping were tested in all data sets. In each scenario data from daughters selected for phenotyping were analysed with a mixed inheritance model with QTL as random effect. Variance components of QTL effects were estimated for five putative QTL positions using restricted maximum likelihood. Gametic relationship matrix between QTL alleles of any two founder haplotypes was computed by two procedures:

**(i)** Linkage association analysis within sire group using a recursive algorithm (Wang *et al.* 1995), for LAC.

**(ii)** LA + LD analysis (combined analysis) across sire family groups (Meuwissen & Goddard 2001), for LDC. In clustering of the haplotypes, all marker loci were used.

The QTL-analysis was carried out using the following linear mixed model:

$$y = \mu + Za + Wq + e$$

where, $\boldsymbol{y}$ is a vector of phenotypes for selected daughters, $\boldsymbol{\mu}$ is overall mean, $\boldsymbol{Z}$ and $\boldsymbol{W}$ are known incidence matrices relating the phenotypes to its polygenic and QTL effects, $\boldsymbol{a}$ is a vector of random additive genetic sire effects, $\boldsymbol{q}$ is a vector of random additive QTL effects and $\boldsymbol{e}$ is a vector of random residual effects. The random variables were assumed to be normally distributed and mutually independent. Specifically a is $N(0, \boldsymbol{A}\sigma^2_{\text{a}})$, $\boldsymbol{q}$ is $N(0, \boldsymbol{H_p}\sigma^2_{\text{q}})$, and $\boldsymbol{e}$ is $N(0, \boldsymbol{I}\sigma^2_{\text{e}})$, where $A$ is the additive relationship matrix, $H_p$ is the IBD matrix that contains the IBD probabilities for the QTL at position $p$, and $I$ is an identity matrix. As sires were assumed to be unrelated, the $\boldsymbol{A}$ matrix simplifies to an identity matrix. The IBD probabilities for the QTL based on the haplotype were calculated using the analytical method of Meuwissen & Goddard (2001).

The DMU program package (Madsen & Jensen 2002) was used for estimation of variance components. An average information restricted maximum likelihood algorithm was used for estimation of (co)variance components and the restricted likelihood is maximized with respect to variance components associated to the random effects (Sørensen *et al.* 2003). The parameters were estimated at the marker brackets (mid points of each marker interval) along the chromosome.

Hypothesis tests to detect QTL were based on the asymptotic distribution of likelihood ratio test (LRT) as:

$$\text{LRT} = -2 \times [\log \text{likelihood}_{(\text{reduced})} - \log \text{likelihood}_{(\text{full})}],$$

where; log likelihood$_{(\text{reduced})}$ is from a model excluding the QTL effect (no QTL model) and log likelihood$_{(\text{full})}$ is the likelihood for a model with a QTL effect, calculated for each bracket. The LRT-statistic has a chi-squared distribution with one degree of freedom. The LRT-statistic thresholds for significant ($p < 0.05$) and highly significant ($p < 0.01$) QTL effects were calculated using the Piepho (2001)which accounts for multiple testing. This was based on using LRT in each putative QTL location, the number of chromosomes, degree of freedom (difference in number of parameters between reduced and full models), and chromosome-wide significance threshold ($\alpha$). Power of detection was assessed as the proportion of the 100 replicates in which a QTL was detected at a given genome-wide significance level ($\alpha$). Accuracy of the position estimate was assessed as the proportion of the 100 replicates in which a QTL was detected in the simulated position $\pm 2$ cM.

## Results

### Linkage analyses sampling criteria

When progeny were selected based on the number of recombinations along the segment using MaxRec and MaxUniRec, no improvement was observed in power

| Strategies | Phenotyping levels (%) | Detection ($\alpha$ = 1%) | Detection ($\alpha$ = 5%) | Position ($\alpha$ = 1%) | Position ($\alpha$ = 5%) |
|---|---|---|---|---|---|
| All[b] | 100 | 69 | 80 | 68 | 80 |
| Random | 30 | 7 | 11 | 6 | 9 |
| | 40 | 9 | 18 | 9 | 15 |
| | 50 | 11 | 29 | 9 | 29 |
| MaxRec | 30 | 2(0.28)[c] | 10(0.91) | 1(0.17) | 9(1.00) |
| | 40 | 9(1.00) | 20(1.11) | 9(1.00) | 20(1.33) |
| | 50 | 17(1.55) | 26(0.90) | 16(1.78) | 26(0.90) |
| MaxUniRec | 30 | 4(0.57) | 11(1.00) | 4(0.67) | 10(1.11) |
| | 40 | 9(1.00) | 14(0.78) | 9(1.00) | 14(0.93) |
| | 50 | 15(1.36) | 30(1.03) | 15(1.67) | 29(1.00) |
| MinRec | 30 | 7(1.00) | 23(2.09) | 7(1.17) | 23(2.56) |
| | 40 | 12(1.33) | 26(1.44) | 10(1.11) | 26(1.73) |
| | 50 | 24(2.18) | 34(1.17) | 24(2.67) | 34(1.17) |

QTL, quantitative trait loci; LA, linkage analyses.
[a]Precision of QTL position shows number of finding around an interval QTL position in simulation out of 100 iterations.
[b]All phenotypes from the offspring were included.
[c]Relative power to random sampling.

**Table 1** Power of QTL detection and precision of position[a] in All (without sampling), Random (sampling strategy at random within family) and three strategies based on LA sampling criteria (MaxRec, MaxUniRec and MinRec) with two levels of error type-I ($\alpha$ = 1 & 5%) using linkage association analysis

to detect QTL and accuracy of position compared to random selection (Table 1) at the same level of phenotyping, e.g. relative power in QTL detection and accuracy of position were 1.36–1.78 at 50% phenotyping and $\alpha$ = 1%. However, using the MinRec which minimizes recombination in paternal haplotypes of the progeny showed that SP was able to increase the power of detection and accuracy of position over random phenotyping (Table 1 and Figure 1), e.g. relative power in QTL detection and accuracy of position were 2.18–2.67 at 50% phenotyping and $\alpha$ = 1%. In this strategy, progeny were sampled with maximum similarity in the paternal haplotypes. The results indicated that sampling a set of non-recombinant offspring has minimized the standard error of the contrast between the sire haplotypes and thereby increased the power of detection in the MinRec compared to sampling of recombinant offspring in other LAC (Table 1). However, LAC showed no consistent increase of relative terms of power compared to random phenotyping with increasing phenotyping levels. Precision of QTL position based on the number of replicates in which the estimated position was within ±2.0 cM from the true position (3.0 cM) was slightly higher in MinRec than MaxRec, MaxUnirec and random SP within family.

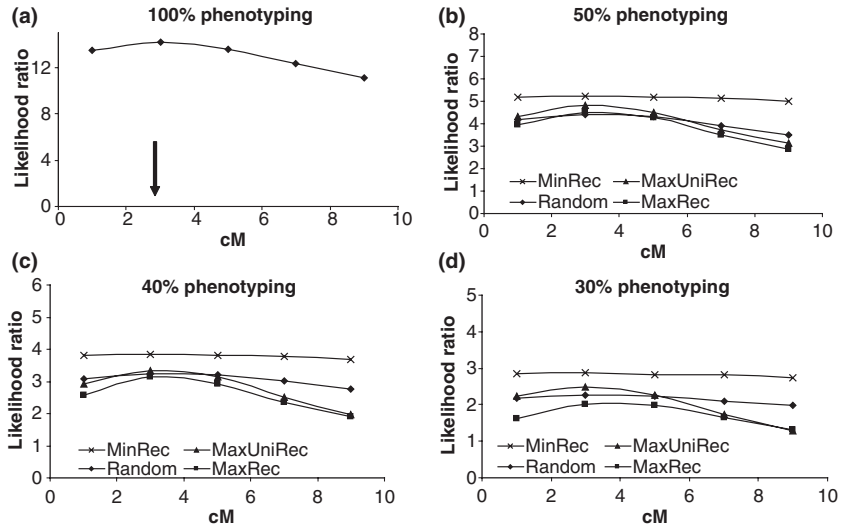


**Figure 1** Map comparisons of different phenotyping strategies using linkage association analysis: (a) all phenotyped individuals (100% phenotyping) with simulated quantitative trait loci position (3.0 cM); (b, c and d) three levels (30, 40 and 50%) of phenotyping based on random (sampling strategy at random within family) and linkage analyses sampling criteria (MaxRec, MaxUniRec and MinRec).

**Table 2** True and mean (±SE) QTL position estimates in cM; gametic QTL (estimated $\sigma^2_{QTL}$), polygenic (estimated $\sigma^2_{Sire}$) and residual variance (estimated $\sigma^2_{error}$) components based on 100 replicates in All (without sampling), Random (sampling strategy at random within family) and three strategies based on LA sampling criteria (MaxRec, MaxUniRec and MinRec) using linkage association analysis

| Strategies | Phenotyping levels (%) | Position (cM) | $\hat{\sigma}^2_{QTL} \pm SE$ | $\hat{\sigma}^2_{Sire} \pm SE$ | $\hat{\sigma}^2_{error} \pm SE$ |
|---|---|---|---|---|---|
| True[a] | | 3.0 | 0.031 | 0.047 | 0.890 |
| All[b] | 100 | 3.90 ± 0.253 | 0.033 ± 0.00150 | 0.049 ± 0.00228 | 0.829 ± 0.0055 |
| Random | 30 | 4.04 ± 0.315 | 0.036 ± 0.00317 | 0.050 ± 0.00324 | 0.816 ± 0.0098 |
| | 40 | 4.10 ± 0.303 | 0.033 ± 0.00254 | 0.050 ± 0.00265 | 0.829 ± 0.0091 |
| | 50 | 3.80 ± 0.274 | 0.032 ± 0.00214 | 0.051 ± 0.00261 | 0.833 ± 0.0078 |
| MaxRec | 30 | 4.18 ± 0.297 | 0.043 ± 0.00401 | 0.041 ± 0.00373 | 0.790 ± 0.0134 |
| | 40 | 4.24 ± 0.278 | 0.044 ± 0.00332 | 0.042 ± 0.00327 | 0.793 ± 0.0111 |
| | 50 | 3.86 ± 0.276 | 0.039 ± 0.00274 | 0.043 ± 0.00288 | 0.806 ± 0.0093 |
| MaxUniRec | 30 | 3.52 ± 0.266 | 0.043 ± 0.00355 | 0.043 ± 0.00334 | 0.794 ± 0.0123 |
| | 40 | 3.86 ± 0.273 | 0.036 ± 0.00263 | 0.046 ± 0.00281 | 0.820 ± 0.0093 |
| | 50 | 3.70 ± 0.261 | 0.036 ± 0.00227 | 0.046 ± 0.00253 | 0.821 ± 0.0081 |
| MinRec | 30 | 4.36 ± 0.340 | 0.036 ± 0.00309 | 0.048 ± 0.00318 | 0.721 ± 0.0111 |
| | 40 | 4.24 ± 0.329 | 0.034 ± 0.00249 | 0.049 ± 0.00282 | 0.826 ± 0.0090 |
| | 50 | 4.00 ± 0.306 | 0.032 ± 0.00212 | 0.048 ± 0.00255 | 0.826 ± 0.0077 |

QTL, quantitative trait loci; LA, linkage analyses.
[a]True value of the parameters used in simulation. Variance component of the QTL is gametic QTL effect.
[b]All phenotypes from the offspring were included.

However, all LAC were unable to estimate QTL position compared to the simulated position (3.0 cM) but they were close to 100% phenotyping (Table 2). In Table 2, the putative QTL position taken as the position of the highest LRT was used in each replicate and then averaged over all replicates. The averages of the likelihood ratios over hundred datasets in Figure 1 indicated that MinRec has higher values across the segment than the other LAC.

Estimated QTL position using all LAC at any of the three phenotyping levels was similar to 100% phenotyping (Table 2). As it might be expected, the sets of progeny selected by the MaxUniRec had fewer recombinations than those selected by the MaxRec (Table 3). With increasing phenotyping levels, the proportion of recombination in recombinant daughters decreased in MaxRec and MaxUniRec. In other word, the MaxUniRec captured 65–70% of the increase in recombinants achieved by the MaxRec.

Gametic QTL variance was estimated as half of the total QTL variance (bi-allelic QTL was assumed). No significant difference was observed between the MinRec and random phenotyping in the estimated gametic QTL effect (Table 2), at the same proportion of phenotyped progeny. The other LAC significantly overestimated the QTL variance at all phenotyping levels. Estimated polygenic variance showed the same pattern and MaxRec and MaxUniRec were always biased and higher than the analysis including all phenotype information. Polygenic variance (from sire model) in this experiment was a quarter of the total additive genetic variance used in the simulation. Therefore, 3/4 of the additive genetic variance was a part of the estimated residual variance component (Table 2).

**Table 3** Mean (±SE) of recombination ratio based on 100 replicates in two strategies based on linkage analyses sampling criteria (MaxRec and MaxUniRec) using linkage association analysis

| Strategies | Phenotyping levels (%) | Recombination ratio |
|---|---|---|
| MaxRec | 10 | 0.9347 ± 0.00894 |
| | 20 | 0.6742 ± 0.01009 |
| | 30 | 0.5731 ± 0.01091 |
| | 40 | 0.5224 ± 0.01139 |
| | 50 | 0.4831 ± 0.01137 |
| MaxUniRec | 10 | 0.7931 ± 0.01036 |
| | 20 | 0.6554 ± 0.01033 |
| | 30 | 0.4924 ± 0.00851 |
| | 40 | 0.3503 ± 0.00553 |
| | 50 | 0.3151 ± 0.00557 |

### Linkage disequilibrium sampling criteria

These criteria were based on LD information across half-sib families to use all available phenotypic and genotypic information and detect QTL using a com-

| Strategies | Phenotyping levels (%) | Detection ($\alpha$ = 1%) | Detection ($\alpha$ = 5%) | Position ($\alpha$ = 1%) | Position ($\alpha$ = 5%) |
|---|---|---|---|---|---|
| All[b] | 100 | 97 | 97 | 97 | 97 |
| Random | 30 | 59 | 77 | 58 | 77 |
| | 40 | 78 | 86 | 77 | 86 |
| | 50 | 91 | 96 | 91 | 96 |
| MaxFre | 30 | 50(0.85)[c] | 64(0.83) | 49(0.84) | 63(0.82) |
| | 40 | 71(0.91) | 87(1.01) | 69(0.89) | 84(0.98) |
| | 50 | 87(0.96) | 91(0.95) | 86(0.95) | 90(0.94) |
| EqHigh | 30 | 57(0.97) | 75(0.97) | 56(0.97) | 75(0.97) |
| | 40 | 81(1.04) | 87(1.01) | 81(1.05) | 87(1.01) |
| | 50 | 88(0.97) | 96(1.00) | 88(0.97) | 94(0.98) |
| Max10 | 30 | 64(1.08) | 81(1.05) | 62(1.07) | 78(1.01) |
| | 40 | 80(1.03) | 89(1.03) | 78(1.01) | 86(1.00) |
| | 50 | 91(1.00) | 96(1.00) | 91(1.00) | 96(1.00) |
| Max50 | 30 | 55(0.93) | 75(0.97) | 54(0.93) | 74(0.96) |
| | 40 | 83(1.06) | 91(1.06) | 82(1.06) | 90(1.07) |
| | 50 | 93(1.02) | 98(1.02) | 92(1.01) | 98(1.02) |

**Table 4** Power of QTL detection and precision of position[a] in All (without sampling), Random (sampling strategy at random across families) and four strategies based on LD sampling criteria (MaxFre, EqHigh,Max10 and Max50) with two levels of error type-I ($\alpha$ = 1 & 5%) using LD + LA analysis

QTL, quantitative trait loci; LA, linkage analyses; LD, linkage disequilibrium.

[a]Precision of QTL position shows number of finding around an interval QTL position in simulation out of 100 iterations.

[b]All phenotypes from the offspring were included.

[c]Relative power compared to random sampling in brackets.

bined LDLA Table 4 presents the power of detecting QTL and precision of estimates of QTL location for the LDC at two levels of alpha (1% and 5%). The QTL precision is given as the number of replicates in which the estimated position is within ±2.0 cM from the true position (3.0 cM). The LDC were more powerful in detecting the QTL than LAC when all progeny were phenotyped (Tables 1 and 4). Additionally, compared to the LAC, using LD information improved power of QTL detection when the percentage of phenotyped individuals increased. The results showed that none of the criteria changed power of QTL detection markedly compared to random phenotyping. Max10 was at least as good as random phenotyping at all levels of phenotyping and both levels of alpha. Precision of estimated QTL locations also increased at increasing levels of phenotyping using the LDC. The Likelihood ratio peaks in Figure 1 were not higher for the LDC when compared to random phenotyping at the higher levels of phenotyping (40% and 50%). However, using LD information to identify the most informative individuals showed similar curves in comparison to random phenotyping except at 30% phenotyping (Figure 2). Both power to detect QTL and precision of position estimates using LDC with maximum similarity among the most frequent clusters (Max10 and Max50) were almost identical at 50% phenotyping

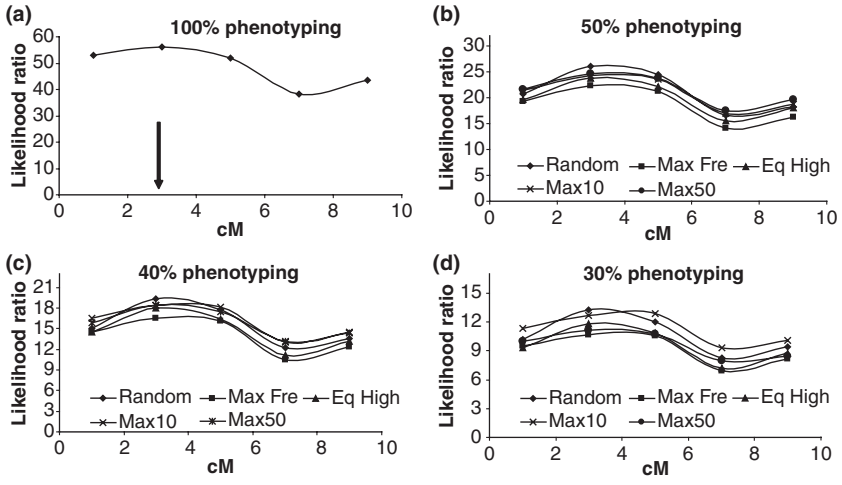


**Figure 2** Map comparisons of different phenotyping strategies using linkage disequilibrium (LD) + linkage analysis: (a) all phenotyped individuals (100% phenotyping) with simulated quantitative trait loci position (3.0cM); (b, c and d) three levels (30, 40 and 50%) of phenotying based on random (sampling strategy at random across families) and LD sampling criteria (MaxFre, EqHigh, Max10 and Max50).

**Table 5** True and mean (±SE) QTL position estimates in cM; gametic QTL (estimated $\sigma^2_{QTL}$), polygenic (estimated $\sigma^2_{Sire}$) and residual variance (estimated $\sigma^2_{error}$) components based on 100 replicates in All (without sampling), Random (sampling strategy at random across families) and four strategies based on LD sampling criteria (MaxFre, EqHigh,Max10 and Max50) using LD + LA analysis

| Strategies | Phenotyping levels (%) | Position (cM) | $\hat{\sigma}^2_{QTL} \pm SE$ | $\hat{\sigma}^2_{Sire} \pm SE$ | $\hat{\sigma}^2_{error} \pm SE$ |
|---|---|---|---|---|---|
| True[a] | | 3.0 | 0.031 | 0.047 | 0.890 |
| All[b] | 100 | 3.46 ± 0.165 | 0.028 ± 0.00097 | 0.052 ± 0.00175 | 0.900 ± 0.0026 |
| Random | 30 | 3.88 ± 0.189 | 0.032 ± 0.00157 | 0.052 ± 0.00255 | 0.902 ± 0.0050 |
| | 40 | 3.94 ± 0.162 | 0.039 ± 0.00142 | 0.051 ± 0.00225 | 0.903 ± 0.0043 |
| | 50 | 3.82 ± 0.164 | 0.032 ± 0.00131 | 0.051 ± 0.00207 | 0.900 ± 0.0038 |
| MaxFre | 30 | 4.28 ± 0.223 | 0.030 ± 0.00180 | 0.055 ± 0.00280 | 0.904 ± 0.0050 |
| | 40 | 4.22 ± 0.209 | 0.030 ± 0.00155 | 0.055 ± 0.00245 | 0.904 ± 0.0043 |
| | 50 | 3.96 ± 0.198 | 0.029 ± 0.00138 | 0.055 ± 0.00224 | 0.903 ± 0.0038 |
| EqHigh | 30 | 4.08 ± 0.221 | 0.031 ± 0.00157 | 0.050 ± 0.00271 | 0.890 ± 0.0049 |
| | 40 | 4.28 ± 0.202 | 0.030 ± 0.00138 | 0.050 ± 0.00232 | 0.894 ± 0.0042 |
| | 50 | 4.06 ± 0.185 | 0.030 ± 0.00126 | 0.051 ± 0.00212 | 0.896 ± 0.0038 |
| Max10 | 30 | 3.96 ± 0.204 | 0.030 ± 0.00184 | 0.054 ± 0.00275 | 0.908 ± 0.0049 |
| | 40 | 4.10 ± 0.208 | 0.031 ± 0.00163 | 0.051 ± 0.00233 | 0.909 ± 0.0043 |
| | 50 | 3.86 ± 0.189 | 0.031 ± 0.00148 | 0.052 ± 0.00217 | 0.907 ± 0.0038 |
| Max50 | 30 | 4.04 ± 0.226 | 0.031 ± 0.00163 | 0.053 ± 0.00324 | 0.898 ± 0.0049 |
| | 40 | 3.86 ± 0.221 | 0.030 ± 0.00143 | 0.054 ± 0.00279 | 0.900 ± 0.0042 |
| | 50 | 3.72 ± 0.194 | 0.029 ± 0.00129 | 0.053 ± 0.00240 | 0.900 ± 0.0038 |

QTL, quantitative trait loci; LA, linkage analyses; LD, linkage disequilibrium.

[a]True value of the parameters used in simulation. Variance component of the QTL is gametic QTL effect.

[b]All phenotypes from the offspring were included.

(Table 4). However, using 30% phenotyping the precision was higher with fewer clusters based on Max10.

In general, the precision of estimated QTL position were not significantly improved as the phenotyping levels increased with the same criterion (Table 5). Using LDC increased the LRT across the chromosomal segment compared to LAC (Figures 1 and 2) because of using all genetic information within and across the half-sib families in the population. The LD sampling criteria were significantly different compared to the estimated QTL position at 100% phenotyping except Max50 at 50% phenotyping. However, at the same level of phenotyping no criteria were significantly different from random phenotyping across families. The results in Table 5 show the MaxFre and EqHigh at low levels of phenotyping (30% and 40%) are further away from the true position than Max10 and Max50. The peaks of LRT in Figure 2 show that all LDC as well as random phenotyping give a wide interval to estimated QTL position (between the markers 2 and 6), regardless of increasing the amount of phenotypes.

Estimates of the QTL, polygenic and residual variance components based on LDC are shown in Table 5. The estimates of variance components using all phenotypic information through LDLA QTL-analysis were considerably closer to the true values compared to LA QTL-analysis (Tables 2 and 5). When 50% of the candidates were randomly phenotyped across the families the estimates of QTL variance were unbiased and close to the simulated value. Results in Table 5 indicate that increasing the percentage of phenotyping to 50%, decreased the estimated QTL variance. However, unbiased QTL estimates were achieved using LDC compared to LAC when a larger segment was used in the simulation (S. Ansari-Mahyari, P. Berg & M.S. Lund, unpublished results). As shown in Table 5, by increasing the proportion of phenotyping in Max10, the estimated QTL effect approached the true value. In contrast to the QTL variance component, the polygenic variance estimate was overestimated at the different levels of phenotyping when LDC were used. Residual variance was similar at all levels of phenotyping. The estimates of polygenic and residual variances using random phenotyping were not significantly different from other strategies when all phenotypes were used.

## Discussion

The current study showed that LDC based on LD information across half-sib families and using LDLA can decrease the required number of phenotyped

individuals compared to LAC and using LA. The results showed that LDC did not significantly improve the power of detecting QTL at the same level of phenotyping relative to random phenotyping. This could be due to the 2.0 cM distance between the markers. Only eight daughters with recombinations are expected among the 80 daughters of each bull in a 10 cM interval. Thus the main difference between the criteria is likely to be whether they sampled daughters with recombinations or not. Among LAC criteria the best was MinRec, sampling daughters without recombinations and thus maximizing information about the contrast between paternal haplotypes. The LAC based on sampling recombinant daughters for genotyping generally performed worse than MinRec and generally not better than random genotyping. They might however be better for larger chromosome segments or larger groups of progeny to select among where more recombinations are expected. For LDC all the criteria aims at utilizing linkage disequilibrium, by sampling daughters with common haplotypes, defined in different ways. This likely favours the selection of daughters without recombinations for phenotyping.

One population with paternal sibs was considered in this study to identify the subsets of phenotyped individuals and applied in detection and localization of a QTL across a chromosomal segment. However, the described criteria can be easily extended to other experimental designs. Casu *et al.* (2003) showed SP through a correlated trait in a daughter design can be useful in QTL detection wherever the main trait is difficult or expensive to record. These studies indirectly used the SP based on information from the correlated traits.

By selecting individuals that have as similar paternal genotype as possible with sire genotypes in each half-sib group using MinRec, selected subsets of offspring providing a better contrast in detection than random sampling. MinRec assigned the individuals in two groups with maximum dissimilarity between different marker genotypes. Therefore, this criterion is better for detection or verification than random phenotyping, however yields poor accuracy to estimate the QTL position compared to the LDLA method. MaxUniRec can be considered when a detected QTL has to be mapped as one of the best criterion for accuracy of QTL location estimates. The higher precision of QTL position estimates using MinRec in comparison to other LAC and random SP in Table 1 was related to higher power of QTL detection. When the number of individuals for phenotyping is limited, LAC by selecting a small number of individuals in each sire family may be improved by selecting a higher fraction of offspring in fewer half-sib families. However, there was little evidence from this simulation study to suggest a considerable improvement when LA based sampling criteria were utilized in selective phenotyping.

The LD information is becoming more popular in QTL fine mapping experiments in outbred population e.g. dairy cattle, because creation of advanced intercrossed lines is nearly impossible in livestock due to time demands, financial constraints and inbreeding depression. If LD exists in the population, one possibility is to use non-random association of alleles at marker brackets for fine mapping (e.g. Bodmer 1986; Xiong & Guo 1997). Different factors can generate LD in a population such as number of founders, population admixture, selection, and mutation. However, modern livestock populations are usually a combination of other smaller populations and therefore, LD could be generated due to differences in the allele frequencies (Nei & Li 1973; Thomson & Klitz 1987). The LDC in this study used the historical ancestral recombinations and contributed substantially both to the probability of detecting the QTL and precision to estimate the QTL position. These criteria were developed based on the LD information in a population which can increase the accuracy of QTL detection (Hayes *et al.* 2006). Using LDLA solely for QTL mapping indicated that fairly moderate QTL effect in simulations (explaining 6.20% of the phenotypic variance) can be detected and positioned. This was not only true with all individuals phenotyped, but also with incomplete phenotype information using LDC.

As results shown, sampling criteria based on LD information across half-sib families and using LDLA can decrease the required number of phenotyped individuals compared to LAC and using LA. Besides, the phenotyped individuals could also be reduced if only LDLA is used rather than LA. The results showed that LDC did not significantly improve the power of detecting QTL at the same level of phenotyping relative to random phenotyping. This could be due to the 2.0 cM distance between the markers. Decreasing the intervals between markers can increase LD information across the families and therefore more benefits of using LDC in comparison to random phenotyping could be expected.

The LAC required large family sizes or prior information about QTL position to identify informative progeny for phenotyping and to successfully detect


© 2009 Blackwell Verlag GmbH 

and position QTL. Therefore, an alternative to LAC in linkage mapping is to use LDC and using both LD and linkage analysis information simultaneously in QTL experiments (Blott *et al.* 2003; Fan *et al.* 2005). The LD information can be used in QTL experiment without any specific family structure within population (Dekkers 2004). In the present study, low levels of phenotypic information have led to overestimated QTL variance components even when selection is based on LDC. Therefore to estimate unbiased genetic (QTL and polygenic) and residual variance components using LDC, it is required to increase the amount of phenotypic information. It should be noted that with increasing population size, the level of phenotyping could be decreased to 30 or even 20% to reach the same power of detection using LDC (S. Ansari-Mahyari, P. Berg & M.S. Lund, unpublished results). In practise LD occurs because most animals in the modern populations have inherited the same piece of chromosome from ancestors perhaps many generations ago (Hayes *et al.* 2006). Therefore, sampling methods based on LD information which could use these common regions (marker haplotypes) might be more reliable to select the informative individuals for phenotyping purposes.

## Conclusion

Generally the sampling criteria tested only resulted in moderate changes in power of QTL detection. Sampling non-recombinant daughters (MinRec) when using LA was at least as good as random phenotyping in all cases studied. The lower resolution for LAC could be expected as the number of recombinant daughters was too low. Among the LDC criteria tested Max10 was the only that was consistently at least as good as random phenotyping. LD information using LDLA analyses was more powerful and it is possible to reach the same power in detection and positioning with less phenotyped progeny if LDC is used instead of LA criteria.

If a dense marker map is available in a region harbouring a QTL to be fine mapped or verified, LDC can be used efficiently with larger difference in power expected than what was found in this study.

## Acknowledgements

This research was supported by a grant from Agriculture Research and Education Organization of Iranian Ministry of Jahad-e-Agriculture. The authors thank Peter Sørensen and Hauke Thomsen for providing their software to estimate IBD probabilities.

## References

Bell P.A., Chaturvedi S., Gelfand C.A., Huang C.Y., Kochersperger M., Kopla R., Modica F., Pohl M., Varde S., Zhao R., Zhao X., Boyce-Jacino M.T. (2002) SNP stream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *BioTechnology (Suppl)*, **32**, S70–S77.

Blott S., Kim J.J., Moisio S., Schmidt-Kuntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J., Georges M., Farnir F., Coppieters W. (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics*, **163**, 253–266.

Bodmer W.F. (1986) Human genetics: the molecular challenge. *Cold Spring Harb. Symp. Quant. Biol.*, **51**(Pt 1), 1–13.

Casu S., Carta A., Elsen J.M. (2003) Strategies to optimize QTL detection designs in dairy sheep populations: the example of the Sarda breed. *Options Mediterraneennes*, **55**, 19–24.

Darvasi A. (1998) Experimental strategies for the genetic dissection of complex traits in animal models. *Nat. Genet.*, **18**, 19–24.

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

Dekkers J.C. (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.*, **82**, 313–328.

Fan R., Spinka C., Jin L., Sun Jung J. (2005) Pedigree linkage disequilibrium mapping of quantitative trait loci. *Eur. J. Hum. Genet.*, **13**, 216–231.

Hayes B.J., Gjuvsland A.B., Omholt S.W. (2006) Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. *Heredity*, **97**, 19–26.

Jannink J.L. (2005) Selective phenotyping to accurately map quantitative trait loci. *Crop Sci.*, **45**, 901–908.

Jin C., Lan L., Attie A.D., Churchill G.A., Bulutuglo D., Yandell B.S. (2004) Selective phenotyping for increased efficiency in genetic mapping studies. *Genetics*, **168**, 2285–2293.

Jobs M., Howell W.M., Stromqvist L., Mayr T., Brookes A.J. (2003) DASH-2: Flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays. *Genome Res.*, **13**, 916–924.

Madsen P., Jensen J. (2002) A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models. Version 6, release 4.4. Danish Institute of Agricultural Sciences, Tjele, Denmark.

Medugorac I., Soller M. (2001) Selective genotyping with a main trait and a correlated trait. *J. Anim. Breed. Genet.*, **118**, 285–295.

Meuwissen T.H.E., Goddard M.E. (2000) Fine mapping of quantitative trait loci using linkage disequilibrium with closely linked marker loci. *Genetics*, **155**, 421–430.

Meuwissen T.H.E., Goddard M.E. (2001) Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.*, **33**, 605–634.

Meuwissen T.H.E., Karlsen A., Lien S., Olsaker I., Goddard M.E. (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, **161**, 373–379.

Nei M., Li W.H. (1973) Linkage disequilibrium in subdivided populations. *Genetics*, **75**, 213–219.

Olsen H.G., Lien S., Svendsen M., Nilsen H., Roseth A., Aasland Opsal M., Meuwissen T.H.E. (2004) Fine mapping milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis. *J. Dairy Sci.*, **87**, 690–698.

Piepho H.P. (2001) A quick method for computing approximate thresholds for quantitative trait loci detection. *Genetics*, **157**, 425–432.

Ansari-Mahyari S., Berg P., Lund M.S. unpublished results.

Searle S.R., Casella G., McCulloch C.E. (1992) Variance Components. John Wiley and Sons, New York.

Sørensen P., Lund M.S., Guldbrandtsen B., Jensen J., Sorensen D. (2003) A comparison of bivariate and univariate QTL mapping in livestock populations. *Genet. Sel. Evol.*, **35**, 605–622.

Thomson G., Klitz W. (1987) Disequilibrium pattern analysis. I. Theory. *Genetics*, **116**, 623–632.

Wang T., Fernando R.L., van der Beek S., Grossman M. (1995) Covariance between relatives for a marked quantitative trait locus. *Genet. Sel. Evol.*, **27**, 251–274.

Wu R., Zeng Z.B. (2001) Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics*, **157**, 899–909.

Wu R., Ma C.X., Casella G. (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics*, **160**, 779–792.

Xiong M., Guo S.W. (1997) Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. *Am. J. Hum. Genet.*, **60**, 1513–1531.


© 2009 Blackwell Verlag GmbH 