Detection of Pleiotropic Effects of Quantitative Trait Loci in Outbred Populations Using Regression Analysis

C. Schrooten*† and H. Bovenhuis*
*Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands and †Holland Genetics, CR Delta, 6802 EB Arnhem, The Netherlands

ABSTRACT

In this paper a method is presented to determine pleiotropic quantitative trait loci (QTL) or closely linked QTL in an outbred population. The method is based on results from single-trait analyses for different traits and is derived for a granddaughter design. The covariance between estimated contrasts of grandsires obtained in single-trait regression analysis is computed. When there is no pleiotropic QTL, the covariance between contrasts depends on the heritabilities of the traits involved, the polygenic and environmental correlation between the traits, the phenotypic standard deviations, the number of sires per grandsire, and the number of daughters per sire. A pleiotropic QTL results in a covariance that deviates from this expected covariance. The deviation depends on the size of the effects on both traits and on the fraction of grandsires heterozygous for the QTL. When analyzing experimental data, the expected covariance and the confidence interval for the expected covariance can be determined by permutation of the data. A covariance outside the confidence interval suggests the presence of a pleiotropic QTL or a closely linked QTL. The method is verified by simulation and illustrated by analyzing an experimental data set on chromosome 6 in dairy cattle.

(Key words: pleiotropy, outbred population, regression analysis, quantitative trait loci)

INTRODUCTION

Genetic improvement of economically important traits in plant and animal species is accomplished by selection based on phenotypic information. Even though the nature of genes influencing these traits is in general not known, selection in this way has been very successful (e.g., Philipsson et al., 1994; Rishell, 1997; Webb, 1998). In the past decade, molecular techniques have become available that allow large-scale genotyping of genetic markers (Gyapay et al., 1996). These markers can be used to map genes affecting quantitative traits (quantitative trait loci, QTL), and many QTL detection studies have been carried out in plant and animal species (e.g., Jansen and Stam, 1994; Georges et al., 1995). This is a first step toward identification and characterization of genes and toward the development of novel selection methods utilizing both phenotypic and genomic information.

Selection for a certain trait can lead to correlated responses for other traits. The main cause for the existence of a genetic correlation between traits in outbred populations is pleiotropy (Falconer and Mackay, 1996), i.e., genes affect more than one trait. Marker-assisted selection (MAS) for a chromosomal region containing genes with pleiotropic effects will thus have consequences for other traits. To avoid negative side effects of selection for certain chromosomal regions or to exploit positive effects of MAS on other traits of interest, the presence of pleiotropic effects should be studied.

The ultimate goal of QTL detection studies is to locate the mutation responsible for the observed effect. This would not only give insight into the biological background of traits but would also allow direct selection for the favorable alleles. However, resolution of QTL detection studies is such that confidence intervals extend over large chromosomal regions. It is expected that regions of this size contain hundreds of genes, as in the human genome (Lander et al., 2001). Further experiments may reduce the length of the confidence intervals, but it is likely that these regions contain many potential candidate genes. Characterization of the QTL based upon their pleiotropic effects will provide additional clues in the identification of candidate genes. Based on the proteins or other metabolites involved in the expression of each of those traits, selection of the most likely candidates from a list of candidate genes in that region is possible. Thus, knowledge about pleio-
tropic effects could speed up the identification of the responsible gene.

Most QTL-detection methods are based on single-trait analysis, either using regression or variance component methods (Fernando and Grossman, 1989; Knott et al., 1994; Grignola et al., 1997). However, methods for multiple-trait analysis have been developed (e.g., Jiang and Zeng, 1995; Weller et al., 1996; Knott and Haley, 2000; Hackett et al., 2001; KoroI et al., 2001). Jiang and Zeng (1995) mentioned three advantages of multiple-trait methods over single-trait methods: multiple-trait methods have higher power, higher precision of parameter estimation, and give potentially more insight into the nature of genetic correlations between different traits. They presented a multiple-trait method based on a maximum-likelihood approach. Weller et al. (1996) used canonical transformation and subsequently single-trait analysis to look at multiple traits. Knott and Haley (2000) developed a multitrait least-squares method for a three-generation pedigree with fixation of the QTL in the grand-parental generation. This method can be extended to other population structures. KoroI et al. (2001) use a transformation of the trait space followed by single-trait analysis and subsequently back transformation. The specific feature of their method is that the multivariate transformations are interval specific. Hackett et al. (2001) extended the regression approach for crosses between lines by Haley and Knott (1992) to a multitrait analysis via multivariate regression and applied this to a doubled haploid population of barley. The application of these methods, however, is not yet common practice. They can present computational difficulties, especially when they would be applied to an outbred population or if the number of traits is large. In the present study, a simple and fast method for detecting pleiotropic QTL or closely linked QTL in an outbred population is presented. The method is based on the covariance between marker contrasts from single-trait regression analysis for different traits. First, the theoretical expectation of this covariance is derived. Second, simulation is used to verify the theoretically expected covariances and to determine the power of detecting pleiotropic effects. Third, the method is applied to data on chromosome 6 in dairy cattle.

MATERIALS AND METHODS

Covariance Between Marker Contrasts

Consider a granddaughter design (Weller et al., 1990) and the following statistical model to analyze the data:

\[
yijkl = gs_i + mj + sk + eijkl, \quad [1]
\]

where \(yijkl\) is the phenotype of animal \(l\), daughter of sire \(k\) that inherited marker allele \(j\) of grandsire \(i\); \(gs_i\) is the effect of grandsire \(i\); \(mj\) is the effect of marker allele \(j\) nested within grandsire \(i\); \(sk\) is the effect of sire \(k\) that received marker allele \(j\) from grandsire \(i\), and \(eijkl\) is the random residual. For each grandsire, two groups of animals can be considered: group 1 consisting of sires that inherited marker allele 1 of the grandsire and group 2 consisting of sires that inherited marker allele 2. The number of sires per grandsire is \(n_s\), and it is assumed that these sires are equally distributed over marker groups 1 and 2. Each sire has \(n_g\) daughters. The contrast between marker allele 1 and marker allele 2 within grandsire \(i\) can be written as \(\bar{y}_{1i} - \bar{y}_{2i}\). The variance of the contrast can be written as

\[
\text{var}(\bar{y}_{1i} - \bar{y}_{2i}) = \frac{4 - h^2}{n_s \times n_g} + \frac{\frac{3}{4}h^2}{n_s} = \sigma^2_{p}, \quad [2]
\]

where \(h^2\) is the heritability and \(\sigma^2_{p}\) is the phenotypic variance of the trait (Weller et al., 1990; van der Beek et al., 1995). Analogously, if we consider two traits, say trait \(v\) and trait \(w\), the covariance between the marker contrasts for trait \(v\) and trait \(w\), for a polygenic situation, is equal to

\[
\text{cov}(\bar{y}_{1i} - \bar{y}_{2i}, \bar{w}_{1i} - \bar{w}_{2i}) = \frac{3r_A h_v h_w + 4r_E \sqrt{1 - h_v^2} \sqrt{1 - h_w^2}}{n_s \times n_g} + \frac{3}{4}r_A h_v h_w \quad [3]
\]

where \(r_A\) is the genetic correlation between trait \(v\) and trait \(w\), \(r_E\) is the environmental correlation between trait \(v\) and trait \(w\), \(h^2_v\) and \(h^2_w\) are heritabilities of trait \(v\) and trait \(w\), respectively, and \(\sigma_{p_v}\) and \(\sigma_{p_w}\) are the phenotypic SD of trait \(v\) and trait \(w\), respectively. When trait \(v\) and trait \(w\) are the same traits, i.e., substituting \(r_A = r_E = 1, h_v = h_w\) and \(\sigma_{p_v} = \sigma_{p_w}\), equation 3 is equal to equation 2. The derivation of equation 3 can be found in the appendix.

As mentioned, the derived covariance of contrasts applies to a polygenic situation. Now consider a bi-allelic QTL, with a fraction \(p_{het}\) of the grandsires heterozygous for the QTL. A fraction \(1 - p_{het}\) is homozygous for the QTL, and in that case the expected contrast between the first and the second grandsire-allele is zero. For the heterozygous grandsires, the contrast between the first and the second grandsire-allele is either positive or negative. In both cases, the absolute size of the contrast is \(\frac{1}{2}a\), where “a” is half the difference between the two homozygous genotypes (Falconer and Mackay, 1996).
The variance of contrasts within each of the three groups is given by equation 2. In the absence of dominance effects, the overall variance of the contrast at the QTL can be written as:

\[
\text{var}(\bar{y}_{1k} - \bar{y}_{2k}) = \frac{3}{2} \sum_{k=1}^{3} \sigma^2_{pk} + \sum_{k=1}^{3} p_k(\mu_k - \mu)^2 = \left\{ \frac{4 - h^2_v}{n_s} + \frac{3h^2_w}{n_g} \right\} \sigma_p^2 + \frac{1}{4} \sigma_{\text{het}}^2 \]

where \( k \) refers to the group (either homozygous, in coupling phase or in repulsion phase), \( p_k \) is the fraction of animals belonging to group \( k \), \( \sigma^2_{pk} \) is the variance of contrasts within group \( k \), \( \mu_k \) is the mean of group \( k \), and \( \mu \) is the overall mean. It should be noted that in equation 4, the heritability and the phenotypic variance are the heritability and the phenotypic variance exclusive of the QTL effect. The variance caused by the QTL effect is covered by the term \( \frac{1}{4} \sigma_{\text{het}}^2 \).

Analogously, the covariance between contrasts for two traits, trait \( v \) and \( w \), can be written as

\[
\text{cov}(\bar{y}_{1v} - \bar{y}_{1w}, \bar{y}_{2v} - \bar{w}_{2w}) = \left\{ \frac{3rAh_vh_w}{n_s} + \frac{4r(1 - h^2_v)(1 - h^2_w)}{n_g} \right\} \sigma_p^2 + \frac{1}{4} \sigma_{\text{het}}^2 \]

where \( a_v \) and \( a_w \) are equal to half the difference between the two homozygous genotypes, for trait \( v \) and trait \( w \), respectively. Heritabilities, variances, and correlations in equation 5 do not include the effect of the QTL. Equation 5 is similar to equation 3, with the addition of the term \( \frac{1}{4} \sigma_{\text{het}}^2 \). This term represents the contribution of the QTL to the covariance. If either \( a_v \) or \( a_w \) is equal to zero, the QTL will not contribute to the covariance of the marker contrasts. Only if both \( a_v \) and \( a_w \) differ from zero, the covariance is affected. The change in covariance due to the QTL depends on the size and the sign of the QTL effect on each trait and on the fraction of grandsires heterozygous for the QTL. The covariance of marker contrasts can be calculated at each location along the chromosome and will deviate for chromosomal regions containing a pleiotropic QTL or closely linked QTL.

**Hypothesis testing.** The null hypothesis (\( H_0 \)) is that the QTL affects only one or none of the two traits under consideration. The alternative hypothesis of a pleiotropic QTL is accepted if the covariance of the contrasts deviates significantly from the covariance under \( H_0 \). The distribution of the covariance under \( H_0 \) is determined by permutation of the data, as first suggested by Churchill and Doerge (1994) for analysis of a single trait. In the current study, pairs of observations on different traits are permuted within families; each pair of trait values is randomly assigned to another member of the same family. In this way, covariances due to unlinked QTL are not affected. After each permutation, the minimum and maximum covariances between marker contrasts on the chromosome are retained. All minimum and maximum values per chromosome are ranked, and the chromosomewise thresholds are determined. In this way, confidence intervals are obtained for the covariance between the contrasts when there is no association between phenotypes and genotypes (i.e., no QTL, equivalent to equation 3).

**Simulated data.** Stochastic simulation was used: 1) to verify the theoretically derived covariances for various situations; 2) to calculate the power of QTL detection by single-trait analysis; and 3) to calculate the power to detect pleiotropic QTL by evaluating the covariance between marker contrasts for two different traits.

Twelve different alternatives were simulated. Each alternative consisted of a granddaughter design with 20 grandsires, 50 sires per grandsire, and 100 daughters per sire. For each grandsire, a chromosome of length 100 cM with 21 equally spaced markers was constructed. Markers were fully informative. Polygenic components for grandsires were generated from normal distributions for trait \( v \) as well as for trait \( w \). Heritability for trait \( v \) was 0.6 and for trait \( w \) the heritability was 0.35. The phenotypic SD for trait \( v \) as well as trait \( w \) was equal to 10. These heritabilities and variances do not include a QTL effect. Polygenic components for sires and their daughters were generated based on half the polygenic component of their sire and a randomly generated term accounting for the contribution of the dam and the Mendelian sampling term. Polygenic components for the second trait were generated conditional on the polygenic components for the first trait, i.e., accounting for the genetic correlation between both traits. A QTL was positioned at 30 cM. The probability that a grandsire was heterozygous for the QTL was 0.5. Marker genotypes for the sires were derived, based on genotypes of their sire and distance between the markers, accounting for recombination. Recombinations were generated using Haldane’s mapping function. Phenotypic records of the daughters for trait \( v \) and trait \( w \) were generated based on the polygenic component, a random environmental component, and the QTL contribution. Environmental components were generated accounting for the environmental correlation between the two traits. The QTL alleles of the daughters were simulated based on the transmitted QTL alleles from
the grandsires and the assumption that both alleles of the biallelic QTL were present in the population at equal frequencies. Genetic correlations were either −0.6 (alternatives 1 to 4, Table 1), zero (alternatives 5 to 8), or +0.6 (alternatives 9 to 12). Phenotypic correlations were assumed equal to the genetic correlations. In the alternatives where a QTL was present, the effect on trait v was 0.3 \sigma_{a_v} where \sigma_{a_v} is the genetic SD for the polygenic effects. This QTL explained 7% of the overall phenotypic variance for trait v. The effect on trait w was either 0 or 0.3 \sigma_{a_w}, explaining 1.6% of the phenotypic variance. When both traits were affected by the QTL, the effect on both traits was either antagonistic (alternatives 3, 7, and 11) or synergistic (alternatives 4, 8, and 12). In alternatives 1, 5, and 9, no QTL was simulated.

For each alternative, 1000 independent datasets were generated. To allow for analysis of phenotypes of sires based on phenotypes of their daughters and analysis at every cM along the chromosome, model (1) was replaced by the multi-marker regression method described by Knott et al. (1994),

\[ Y_{ij} = \mu + g_{si} + b_{ik}X_{ijk} + e_{ij}, \]

where \( Y_{ij} \) = daughter yield deviations (VanRaden and Wiggans, 1991) of sire j, son of grandsire i, based on 100 daughters, \( \mu \) = overall mean, \( g_{si} \) = fixed effect of grandsire i, \( b_{ik} \) = regression coefficient for grandsire i at position k on the chromosome, \( X_{ijk} \) = probability that sire j receives a chromosomal segment from grandsire i at position k, and \( e_{ij} \) = random residual.

In each family, estimates of the contrast between allelic effects, the b-values in equation 6, were obtained at every centimorgan along the chromosome, for both traits. Covariances based on the contrasts of the 20 grandsire families in the analysis were computed at every centimorgan. To determine whether results from simulation were in agreement with the derived equations, the average covariance at the simulated location of the QTL, based on 1000 datasets, was compared to the covariance according to equation 5. Besides, the average covariance after permutation was compared to the covariance according to equation 3, with parameters adjusted to account for additional variance caused by the QTL. Adjusted parameters are in Table 2.

To determine the sensitivity of the method for the number of heterozygous grandsires, alternatives 2 and 3 were also studied with 10 and 5 families instead of 20. For the same reason, alternatives 2 and 3 were also studied with 20 families and 25% heterozygous grandsires instead of 50%.

**Power.** To test the hypothesis that there is no QTL affecting a certain trait, 2000 permutations were applied to each data set. A QTL was considered significant when the test statistic exceeded the 5% chromosome-wise threshold. The probability of detecting a QTL for individual traits was obtained by determining the

Table 1. Probability (Prob.) to detect QTL and pleiotropic QTL, covariances before and after permutation and confidence intervals in each alternative.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>rA, rP, a, a</th>
<th>Prob.</th>
<th>QTL, trait v</th>
<th>Prob.</th>
<th>QTL, trait w</th>
<th>Expected covariance at QTL</th>
<th>Covariance after permutation</th>
<th>Expected covariance after permutation</th>
<th>Covariance after permutation</th>
<th>Lower threshold</th>
<th>Upper threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−0.6 0</td>
<td>0</td>
<td>0.06 0.06</td>
<td>0.05</td>
<td>−0.45</td>
<td>−0.46</td>
<td>−0.45</td>
<td>−0.45</td>
<td>−0.48</td>
<td>−1.25</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>−0.6 0.5</td>
<td>0</td>
<td>0.87 0.06</td>
<td>0.20</td>
<td>−0.45</td>
<td>−0.47</td>
<td>−0.45</td>
<td>−0.48</td>
<td>−1.28</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>−0.6 0.5 0</td>
<td>0.3</td>
<td>0.86 0.32</td>
<td>0.91</td>
<td>0.40</td>
<td>0.44</td>
<td>−0.40</td>
<td>−0.42</td>
<td>−1.20</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>−0.6 0.5 0</td>
<td>−0.3</td>
<td>0.87 0.32</td>
<td>0.61</td>
<td>−1.31</td>
<td>−1.35</td>
<td>−0.51</td>
<td>−0.54</td>
<td>−1.38</td>
<td>−0.01</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.04 0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>−0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.5</td>
<td>0.88 0.23</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>−0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.5</td>
<td>0.3 0.86</td>
<td>0.81</td>
<td>0.86</td>
<td>0.89</td>
<td>0.05</td>
<td>0.06</td>
<td>−0.53</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.5</td>
<td>−0.3 0.88</td>
<td>0.34</td>
<td>−0.86</td>
<td>−0.91</td>
<td>−0.05</td>
<td>−0.06</td>
<td>−0.68</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.6</td>
<td>0</td>
<td>0.04 0.05</td>
<td>0.05</td>
<td>0.45</td>
<td>0.46</td>
<td>0.45</td>
<td>0.48</td>
<td>−0.01</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.6 0.5</td>
<td>0</td>
<td>0.84 0.21</td>
<td>0.45</td>
<td>0.46</td>
<td>0.45</td>
<td>0.45</td>
<td>0.48</td>
<td>−0.04</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.6 0.5 0</td>
<td>0.3</td>
<td>0.87 0.35</td>
<td>0.61</td>
<td>1.31</td>
<td>1.34</td>
<td>0.51</td>
<td>0.54</td>
<td>0.01</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.6 0.5 0</td>
<td>−0.3</td>
<td>0.89 0.29</td>
<td>0.90</td>
<td>−0.40</td>
<td>−0.43</td>
<td>0.40</td>
<td>0.43</td>
<td>−0.10</td>
<td>1.21</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Alternatives are characterized by size of the simulated effects for trait v and trait w (a_v and a_w, respectively), expressed in genetic SD units, and by the phenotypic (r_p) and genetic correlation (r_A).

\(^2\)Type 1 error in alternatives 1, 5 and 9, power in other alternatives.

\(^3\)Type 1 error in alternatives 1, 2, 5, 6, 9 and 10, power in other alternatives.

\(^4\)Determined as the sum of the fraction above the upper threshold and the fraction below the lower threshold.

\(^5\)Calculated using Equation 5.

\(^6\)Average of 1,000 simulated datasets.

\(^7\)Calculated using Equation 3, with parameters adjusted to account for additional variance and covariance caused by the quantitative trait locus (QTL). Adjusted parameters are in Table 2.

\(^8\)Average of 1,000 simulated datasets and 2,000 permutations for each dataset. Presented thresholds are for 95% confidence interval.
Table 2. Adjusted parameters to be used in Equation 3 when computing the covariance between contrasts after permutation when a QTL is present; parameters need to be adjusted for additional variance and covariance caused by the QTL.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>r_A, r_P^2</th>
<th>a_v</th>
<th>a_w</th>
<th>r_A^3</th>
<th>r_P^3</th>
<th>h_v^3</th>
<th>h_w^3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.6</td>
<td>0</td>
<td>0</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.35</td>
</tr>
<tr>
<td>2</td>
<td>-0.6</td>
<td>0.5</td>
<td>0</td>
<td>-0.57</td>
<td>-0.58</td>
<td>0.63</td>
<td>0.35</td>
</tr>
<tr>
<td>3</td>
<td>-0.6</td>
<td>0.5</td>
<td>0.3</td>
<td>-0.48</td>
<td>-0.54</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>4</td>
<td>-0.6</td>
<td>0.5</td>
<td>0.3</td>
<td>-0.62</td>
<td>-0.61</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.60</td>
<td>0.35</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.63</td>
<td>0.35</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0.5</td>
<td>0.3</td>
<td>0.07</td>
<td>0.03</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.60</td>
<td>0.30</td>
</tr>
<tr>
<td>9</td>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.35</td>
</tr>
<tr>
<td>10</td>
<td>0.6</td>
<td>0.5</td>
<td>0</td>
<td>0.57</td>
<td>0.58</td>
<td>0.63</td>
<td>0.35</td>
</tr>
<tr>
<td>11</td>
<td>0.6</td>
<td>0.5</td>
<td>0.3</td>
<td>0.62</td>
<td>0.61</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>12</td>
<td>0.6</td>
<td>0.5</td>
<td>-0.3</td>
<td>0.48</td>
<td>0.54</td>
<td>0.63</td>
<td>0.36</td>
</tr>
</tbody>
</table>

1 Alternatives are characterized by size of the simulated effects for trait v and trait w (a_v and a_w, respectively), expressed in genetic SD units, and by the phenotypic (r_P) and genetic correlation (r_A).
2 Genetic and phenotypic correlation based on polygenic components, assuming no quantitative trait locus (QTL).
3 Genetic and phenotypic correlation and heritabilities for trait v and trait w corrected for additional variance and covariance caused by the QTL.

The method was illustrated using data from a granddaughter design in dairy cattle, consisting of 20 grandsires and 833 sires. The number of sires per grandsire varied from 11 to 147. Genotypes for 29 microsatellite markers on chromosome 6 were available. Pleiotropic QTL effects were studied considering two traits: milk yield and protein percentage. In this data, a QTL affecting protein percentage was identified previously (Spelman et al., 1996). Further, there was a suggestive QTL for milk yield located in the same chromosomal region as the QTL affecting protein percentage, and a second suggestive QTL for milk yield more than 40 cM away from the first QTL.

Estimated breeding values for protein percentage and milk quantity were converted to daughter yield deviations by deregression, and subsequently analyzed using equation 6 (Spelman et al., 1996). The 95% confidence interval for the covariance under the null hypothesis was based on 15,000 permutations. Variation in number of daughters per sire was accounted for by applying a weighted regression when using equation 6, as described by Spelman et al. (1996). Variation in the number of sires per family was accounted for by weighing the contrasts using the inverse of the standard error, i.e., contrasts with large standard error had less weight in computation of the covariance.

**RESULTS**

**Simulation.** In the simulated alternatives, the covariance between contrasts at the QTL was in agreement with the expected covariance. The difference between expected and realized covariances was never larger than 0.05 (Table 1). In Figure 1, the pattern of the covariance in alternative 3, averaging over 1000 independent replicates, is presented. The average covariance after permutation and the 95% confidence interval, derived from 2000 permutations of each of 1000 datasets, are indicated as well. The largest difference between realized and expected covariance was found at the location of the QTL. The difference between the covariance and the expected covariance decreased with increasing distance to the QTL.

**Power of QTL detection.** Table 1 shows the power of QTL detection, for each simulated alternative and each trait. Power to detect the QTL for trait v, with an effect of 0.5 σ_v, ranged from 0.84 to 0.89, with an average over relevant alternatives of 0.87. In 90% of the cases with a significant QTL affecting trait v, the location of the highest test statistic was in the interval between 20 and 40 cM, containing the simulated QTL at 30 cM.
Figure 1. Average covariance and significance thresholds for the 95% confidence interval under the H₀-hypothesis, based on 1,000 replicates of alternative 3, with genetic and phenotypic correlation equal to −0.6. A quantitative trait locus QTL with effect 0.5 \( \sigma_a \) on trait \( v \) and 0.3 \( \sigma_a \) on trait \( w \) was located at 30 cM. Confidence interval (95%, chromosomewise) and expected covariance are based on 2,000 permutations of each dataset and averaged. Position of the markers indicated on the horizontal axis (◆).

The power to detect the QTL for trait \( w \), with an effect of 0.3 \( \sigma_a \), ranged from 0.29 to 0.35, with an average of 0.32. The QTL was positioned in the interval of the simulated QTL, i.e., between 20 and 40 cM, in 66% of the cases in which there was significant evidence for the presence of a QTL. The probability of detecting a QTL when no QTL was simulated (i.e., the type I error), ranged from 0.04 to 0.06, with an average of 0.05 for trait \( v \) as well as for trait \( w \), across all relevant alternatives.

**Power to detect pleiotropic QTL.** In the alternatives where a pleiotropic QTL was simulated, the power to detect a QTL with a pleiotropic effect ranged from 0.61 to 0.91 (Table 1). The power was higher if the covariance due to the QTL and the polygenic and environmental covariance had opposite sign, as in alternatives 3 and 12. If no QTL was simulated, there was a probability of 0.05 to detect a QTL with a pleiotropic effect, which is equal to the expected probability. If a QTL with only an effect on trait \( v \) was simulated, the probability of detecting a pleiotropic QTL was 0.20 (alternative 2) to 0.23 (alternative 6), which is higher than the expected probability of 0.05 based on permutation. Threshold levels to determine the power were based on permutations. The average covariance after permutation was in agreement with the expected covariance according to equation 3, with parameters adjusted for additional variance caused by the QTL (Table 2). Size of the confidence intervals ranged from 1.12 in alternative 5 to 1.37 in alternatives 4 and 11. Confidence intervals were larger in alternatives where a QTL-effect was simulated. Results in alternatives 1, 2, 3, and 4 were similar to the results in the corresponding alternatives 9, 10, 12 and 11, respectively.

**Number of heterozygous grandsires.** Alternatives 2 and 3 were simulated with 10 and 5 families instead of 20, and with 20 families and 25% heterozygous grandsires instead of 50%. These simulations showed, that the method can also be applied if the number of grandsires is less than 20, although the power is reduced (results not shown).

**Experimental data.** Single-trait analyses revealed a QTL affecting protein percentage and possibly kilograms of milk in the interval between zero and 50 cM. In addition, a suggestive QTL affecting kilograms of milk was found between 75 and 113 cM. The pattern of the covariance between contrasts for kilograms of milk and protein percentage along chromosome 6 is given in Figure 2. At the beginning and the end of the chromosome, the covariance significantly deviates from the expected covariance under the H₀-hypothesis. Figure 2 shows two confidence intervals: one interval is determined by retaining the maximum and minimum value of the covariance on the chromosome in each permutation (chromosomewise thresholds); the other interval is determined by retaining the maximum and
Figure 2. Covariance between contrasts for milk yield and protein percentage on chromosome 6. Confidence interval (95%) and expected covariance under the H0-hypothesis (no pleiotropic quantitative trait locus [QTL]), are based on 15,000 permutations of the dataset. Confidence intervals based on chromosomewise and nominal values are indicated in the figure. Expected covariance based on chromosomewise values is indicated by a solid line, expected covariance based on nominal values (based on permutation at each location), is indicated by (++) . Position of the markers indicated on the horizontal axis (◆).

minimum values of the covariance at each location separately (nominal thresholds). In the latter case, confidence intervals are not constant along the chromosome.

**DISCUSSION**

The presented method offers an easy way to determine pleiotropic effects of a QTL, as the method makes use of results from single-trait regression analysis. In this paper, the method is derived and is illustrated for a granddaughter design, but the method can also be derived for a two-generation design like a daughter design. Estimated contrasts for grandsires from the regression analysis are used to compute the covariance between these contrasts for different traits at each chromosomal location. Permutation is carried out to determine confidence intervals for the covariance, in order to detect significant deviations from the expected covariance under the null hypothesis of no QTL or a QTL affecting only one trait. The expected covariance in this situation is not necessarily 0, but a function of the genetic and environmental correlation, the heritabilities, the phenotypic variances, and the number of sires and granddaughters involved (equation 3).

**Power and type I error.** In single-trait analysis, power to detect a QTL with an effect of 0.5 $\sigma_a$ ranged from 0.84 to 0.89. When applying power calculations as presented by Weller et al. (1990), power was 0.93. In alternatives with a QTL effect on trait w of 0.3 $\sigma_a$, power ranged from 0.29 to 0.35. Calculation of the power using Weller et al. (1990) gave a power of 0.43. Considering the assumptions underlying the power calculations by Weller et al. (1990), it can be concluded that power based on simulations was in agreement with expectations.

The power to detect a pleiotropic QTL effect was high for the simulated alternatives. If the QTL effect on trait v was 0.5 $\sigma_a$ and the effect on trait w was 0.3 $\sigma_a$, the power to detect the pleiotropic effect ranged from 0.61 (alternatives 4 and 11) to 0.91 (alternative 3). For the alternatives where the QTL had opposite effects on both traits, power ranged from 0.61 (alternative 4) to 0.90 (alternative 12). The highest power was obtained in the alternatives where the covariance due to the QTL and the covariance based on residual genetic and environmental effects had different signs. This indicates that QTL that “break” the overall genetic correlation can be detected more easily. This was also noted by Korol et al. (1995).

It should be noted, however, that the type I error was higher than expected: Type I errors ranged from 0.20 to 0.23. In alternative 2, for example, a Type I error of
20% was obtained, where 5% was expected, based on permutation. When no QTL effect was simulated, i.e., in alternatives 1, 5 and 9, type I errors derived from permutation agreed with expected values. The discrepancy between the expected and observed type I error in other alternatives is caused by the fact that permutation does not result in the correct distribution under the null hypothesis in case there is a QTL affecting only one of the traits. This is illustrated in Figure 3 for alternative 2. Figure 3 shows the actual distribution of the covariance (average of 1000 independent datasets) and the distribution of the covariance based on permutation (average of 2000 permutations of these 1000 datasets). Covariances are shown at the QTL. Although expectations of both distributions are equal, the variance of the distribution based on permutations is lower. Consequently, the confidence interval for the covariance based on permutations will be underestimated, resulting in more-than-expected false positives.

Obviously, when analyzing experimental data, this will also result in more false positives than expected based on type I errors set by permutation. Simulation showed that the number of false-positive pleiotropic QTL increases with increasing effect of the QTL. When the QTL effect was 0.2 $\sigma_a$, the observed type I error was 0.12 and 0.20, respectively. To obtain 5% false positives for a QTL effect of 0.5 $\sigma_a$, the type I error based on permutation should be set to 0.2%. Thus, using lower type I errors in constructing confidence intervals can reduce the number of false positives to an acceptable level. This will, however, also reduce the power. To indentify false-positive results, the results from the underlying single-trait analyses could also be considered.

**Experimental data.** Based on the pattern of the covariance and the confidence interval in Figure 2, it can be concluded that chromosome 6 contains two pleiotropic QTL, affecting both kilograms of milk and protein percentage. Selection for alleles that increase milk quantity results in a decrease of protein content. This information should be taken into account when deciding whether or not to apply MAS for these QTL.

When confidence intervals for the experimental data were based on nominal values, it appeared that they were not constant along the chromosome. Deviations were observed especially at the beginning and the end of the chromosome. Simulation showed that this effect can be reproduced by simulating data with a low marker density at the beginning and the end of the chromosome. Simulated data containing markers with low information content at the beginning and the end of the
Figure 4. Correlation between contrasts in alternative 2 ($r_a$ and $r_p$ equal to -0.6), alternative 6 ($r_a$ and $r_p$ equal to 0.0), and alternative 10 ($r_a$ and $r_p$ equal to +0.6), based on 1,000 replicates of each alternative. The effect of the quantitative trait locus (QTL) on trait $v$ was 0.5 $\sigma_v$ in all three alternatives. Position of the markers indicated on the horizontal axis ($\blacklozenge$).

chromosome gave similar results. When chromosomewise thresholds are determined for these situations, thresholds will predominantly be influenced by values at the beginning and the end of the chromosome. Therefore, thresholds can be too strict for other regions of the chromosome, resulting in lower power to detect pleiotropic QTL in those regions. This problem can be overcome by adjusting thresholds based on nominal values for repeated testing along the chromosome, i.e., using stringent nominal values.

**Biallelic vs. multi-allelic QTL.** Both equation 4 and the simulations were based on the assumption that there is an additive biallelic QTL. The contribution of grandsires heterozygous for the QTL to the covariance at the QTL is $\frac{1}{4} * a_v * a_w$. Each of the two possible heterozygous grandsires, in coupling or in repulsion phase, contributes equally to the covariance. In reality, QTL can be multi-allelic, or actually consist of two closely linked biallelic QTL, which together act as a multi-allelic QTL. For this situation the contribution of a heterozygous grandsire to the covariance can have a range of values. Depending on the alleles present and their effects, the contribution can be positive, negative, or zero. Summing over all grandsires, the overall effect on the covariance could be close to the expected covariance under the null hypothesis. As a consequence, the pleiotropic QTL might not be detected in all multi-allelic situations.

**Pleiotropic vs. linked QTL.** Instead of considering a QTL with a pleiotropic effect, one could also consider two linked QTL that are involved in the expression of two traits. This can also cause a significant covariance between the estimated contrasts. The probability of detecting a significant covariance between contrasts decreases with increasing distance between the linked QTL. The effect of distance between linked QTL on the covariance was not studied here. Notably, however, even fairly well-spaced QTL could cause a significant covariance, provided the QTL effect on each trait is large enough. The method presented here does not distinguish between pleiotropy and linkage, but if the goal is to look at the possible consequences of MAS for other traits, it is less relevant if the QTL is a pleiotropic QTL or if there are two closely linked QTL.

**Covariance vs. correlation.** In this study, the covariance between contrasts was used to study QTL with an effect on two traits. It is also possible to look at the correlation between contrasts. Combining equations 3 and 4, it follows that if there is a QTL affecting one trait, the covariance is equal to the covariance when there is no QTL, but the variance of the contrasts increases. As a result, the absolute value of the correlation
between contrasts will be lower than in a situation without a QTL (Figure 4). Thus, permutation will not give insight if the deviation of the correlation is due to a QTL affecting two traits, or due to a QTL affecting only one trait. When looking at the covariance, a significantly deviating covariance can be detected by permutation, provided that the right type I errors are used.

CONCLUSIONS

This study has shown that the covariance between contrasts from separate single-trait regression analyses can be used to identify pleiotropic or closely linked QTL, i.e., QTL or linked QTL that act on more than one trait. The presented method gives insight in potential benefits or drawbacks from MAS due to the effect of MAS on nontarget traits. Results from simulations showed that stringent Type I errors should be applied to reduce the number of false-positive results. Application of the method to chromosome 6 indicated two QTL affecting protein percentage as well as milk yield. When MAS is used to increase the level of one of these traits, the level of the other trait will decrease.

ACKNOWLEDGMENTS

We thank Johan van Arendonk, Marco Bink, and Sijne van der Beek for their comments in preparing the manuscript. We appreciate the financial support by Holland Genetics.

REFERENCES


APPENDIX

Covariance Between Contrasts for Two Traits from Regression Analysis of a Granddaughter Design, in the Absence of a QTL

Consider a granddaughter design with the following analysis model

\[ y_{ijkl} = g_{si} + m_{ij} + s_{ijk} + e_{ijkl} \]  

where \( y_{ijkl} \) is the phenotype of animal \( l \), daughter of sire \( k \) that inherited marker allele \( j \) of grandsire \( i \); \( g_{si} \) is the effect of grandsire \( i \); \( m_{ij} \) is the effect of marker allele \( j \) nested within grandsire \( i \); \( s_{ijk} \) is the effect of sire \( k \) that received marker allele \( j \) from grandsire \( i \), and \( e_{ijkl} \) is the random residual.

Journal of Dairy Science Vol. 85, No. 12, 2002
Suppose that the number of sires per grandsire is \( n_s \) and that each grandsire has \( n_g \) granddaughters per sire. With equal distribution of sires over marker alleles, the number of sires per grandsire per marker allele is \( \frac{1}{2} n_s \). When considering two traits, \( v \) and \( w \), the covariance of the contrast of marker group means can be written as:

\[
\text{cov}(\overline{v_{i1.} - v_{i2.}}, \overline{w_{i1.} - w_{i2.}}) = \frac{1}{(\frac{1}{2} n_s n_g)^2} \left( \frac{1}{2} n_s n_g \cdot \text{cov}(v_{i111}, w_{i111}) + \frac{1}{2} n_s n_g(n_g - 1) \cdot \text{cov}(v_{i111}, w_{i112}) + \frac{1}{4} n_s n_g n_g \cdot \text{cov}(v_{i111}, w_{i121}) \right) + \text{cov}(\overline{v_{i2.} - v_{i1.}}, \overline{w_{i2.} - w_{i1.}}) \tag{2}
\]

Group \( i1 \) (same grandsire and same marker) consists of three different categories when considering variances and covariances: animals themselves, half sibs (same sire, different dam), and animals only related through their paternal grandsire.

\[
\text{cov}(v_{i111}, w_{i111}) = \text{covariance between records on animal itself} = r_p \sigma_v \sigma_p w = \left( r_A h_v h_w + r_E \sqrt{1 - h_v^2} \sqrt{1 - h_w^2} \right) * \sigma_v \sigma_p w \tag{4}
\]

\[
\text{cov}(v_{i111}, w_{i112}) = \text{covariance between records on half sibs} = \frac{1}{4} r_A h_v h_w \sigma_v \sigma_p w \tag{5}
\]

\[
\text{cov}(v_{i111}, w_{i121}) = \text{covariance between records on animals with same grandsire, different sire} = \frac{1}{16} r_A h_v h_w \sigma_v \sigma_p w \tag{6}
\]

\[
\text{cov}(\overline{v_{i1.} - v_{i2.}}, \overline{w_{i1.} - w_{i2.}}) = \sigma_p \sigma_p w \left( \frac{1}{2} n_s n_g (r_A h_v h_w + r_E \sqrt{1 - h_v^2} \sqrt{1 - h_w^2}) \right) + \frac{1}{2} n_s n_g (n_g - 1) \cdot \frac{1}{4} r_A h_v h_w \sigma_v \sigma_p w \tag{7}
\]


\[
\text{cov}(\overline{v_{i1.} - v_{i2.}}, \overline{w_{i1.} - w_{i2.}}) = \frac{1}{2} n_s n_g \left( \frac{4 r_A h_v h_w + 4 r_E \sqrt{1 - h_v^2} \sqrt{1 - h_w^2}}{n_s n_g} \right) + \frac{1}{2} r_A h_v h_w \sigma_v \sigma_p w \tag{Equation 3}
\]

\[
\text{cov}(\overline{v_{i2.} - v_{i1.}}, \overline{w_{i2.} - w_{i1.}}) = \text{cov}(\overline{v_{i1.} - v_{i2.}}, \overline{w_{i2.} - w_{i1.}})
\]

\[
\text{cov}(\overline{v_{i1.} - v_{i2.}}, \overline{w_{i1.} - w_{i2.}}) = \frac{3 r_A h_v h_w + 4 r_E \sqrt{1 - h_v^2} \sqrt{1 - h_w^2}}{n_s n_g} + \frac{3 r_A h_v h_w}{n_s n_g} \sigma_v \sigma_p w
\]