The number of loci that affect milk production traits in dairy cattle.

Amanda Jane Chamberlain*,†
Helen Clare McPartlan*
Michael Edward Goddard*,†

* Animal Genetics and Genomics Platform, Department of Primary Industries Victoria, Victoria, Australia, 3049.
† Land and Food Resources, University of Melbourne; Parkville, Victoria, Australia, 3052.
Estimating the number of loci.

quantitative trait loci/ milk production/ quantitative trait/ gene mapping/
false discovery rate

Amanda Chamberlain
Department of Primary Industries Victoria
475 Mickleham Rd
Attwood, 3049
Australia
Ph: +61 3 9217 4195
Fax: +61 3 9217 4299

amanda.chamberlain@dpi.vic.gov.au
ABSTRACT

We have used the results of an experiment mapping quantitative trait loci (QTL) affecting milk yield and composition to estimate the total number of QTL affecting these traits. We did this by estimating the number of segregating QTL within a half-sib daughter design using logic similar to that used to estimate the “false discovery rate” (FDR). In a half-sib daughter design with 6 sire families we estimate that the average sire was heterozygous for approximately five QTL per trait. Also, in most cases only one sire was heterozygous for any one QTL, therefore at least 30 QTL were likely to be segregating for these milk production traits in this Holstein population.
Milk production and composition are classical quantitative traits, often described by the infinitesimal model, which assumes that the number of loci is infinitely large, each with an infinitely small effect (Lynch and Walsh 1998). However, QTL mapping experiments have shown that there are not really an infinite number of loci and that some genes have a moderate effect on quantitative traits, i.e., DGAT1 on chromosome 14 (Grisart et al. 2004) and GHR on chromosome 20 (Blott et al. 2003). In fact, Thoday and Thompson (Thoday and Thompson 1976) showed that a normal distribution of phenotypes in a segregating population can be the result of as few as three independently segregating loci. The infinitesimal model produces good predictions of short-term selection response, but for prediction of long-term response, and especially for predicting the value of marker assisted selection, it would be useful to know the number of real QTL underlying a typical quantitative trait. This would improve our knowledge of the genetics of quantitative traits, which would improve our prediction of the results of natural and artificial selection, as well as planning for marker assisted selection.

The classical method of Wright (1968) for estimating the number of genes underlying a quantitative trait is simple and has been widely used. However due to the failure of many of it’s assumptions it is seriously biased. Lande (1981), Zeng et al. (1990) and Zeng (1992) have all applied variations of
this estimator to predict the number of loci having an effect on quantitative traits. However, they have all failed to correct all the incorrect assumptions of Wright. Because of this Wright’s estimator has been considered unsuitable for general use, though it may be used to give an estimate of the minimum number of genes. The availability of molecular data has seen new methods for estimating the number of genes or QTL. Otto and Jones (2000) described a QTL based estimator for the total number of loci. Other groups have estimated the number of QTL from looking at a number of QTL mapping studies (Beavis 1994; Doebley and Stec 1991; Doebley et al. 1990). All of these studies used only significant QTL, and did not account for false positives or false negatives.

When many hypotheses are tested, as is the case in whole genome scans for QTL, the concept of false discovery rate (FDR) has been introduced as an alternative to classical significance testing (Benjamini and Hochberg 1995; Weller et al. 1997). The logic behind the FDR provides an estimate of the total number of genuinely false null hypotheses, which is the number of real QTL. In this paper the logic behind the FDR was used, not to determine which QTL were significant, but to estimate across a whole genome-scan, the number of QTL that were segregating per trait. We did this by estimating the number of segregating QTL within a half-sib daughter design.
MATERIALS AND METHODS

A selective genotyping experiment consisting of six sires was used, each with approximately 100 high and 100 low Australian Selection Index (ASI) daughters. ASI is an economic index of milk, fat and protein yields. These animals were genotyped for 130 microsatellite markers distributed across 21 chromosomes. Linkage analysis was performed using QTL express (SEATON et al. 2002). Phenotypes used in the linkage analysis, provided by the Australian Dairy Herd Improvement Scheme, were deregressed Australian Breeding Values (ABVs) for protein yield, fat yield, milk yield, protein percentage, fat percentage, fat percentage corrected for protein percentage (F%-P%) and ASI, where F%-P% = F%-1.43P% and ASI = (3.8*protein) + (0.9*fat) – (0.048*milk). Analyses were fixed at marker midpoints throughout the genome, only positions greater than 15cM apart were chosen for this analysis. The final data set consisted of a total of 89 chromosomal positions, for between one and six sires, a total of 410 QTL tests. A total of seven traits were tested, resulting in 2870 data points. Chromosome, trait, family, and estimate of QTL effect, along with its standard error and absolute t-value, were extracted from the raw data for all of these positions. The absolute t-values were converted to a p-value.
Three analyses were performed on this data set. First, the number of heterozygous QTL per sire per trait was estimated. Second, the number of heterozygous QTL per sire per trait was corrected for artefacts caused by linkage between adjacent chromosome segments. Third, the number of heterozygous sires for any given QTL was estimated. In each case, the data set is modelled in terms of a parameter, and maximum likelihood (ML) used to estimate this parameter. In some cases the ML estimate was obtained by equating the expected and observed proportions and in other cases by numerically maximising the likelihood in Excel™ (Microsoft®).

**Estimation of the proportion of heterozygous QTL**

Because many significance tests are performed, we expect to find some false positive QTL. We can correct for these when estimating the total number of real QTL, if the QTL that are really heterozygous are assumed to be always significant. Then, the expected proportion of marker brackets that contain an apparently significant QTL \((x)\) can be predicted as

\[
x = q + (1 - q)p
\]

where \(q\) is the proportion of marker brackets that contain a heterozygous QTL, all of which are significant, and \(1-q\) is the proportion of marker brackets that do not contain a heterozygous QTL, of which a proportion \(p\) are significant (false positives), when a point-wise significance level of \(p\) was used. Therefore this calculation made the assumption that anything that
was not significant was in fact a true case of there being no QTL, but anything that was significant might have been a false positive.

This expected proportion of significant marker brackets was equated to the observed proportion \( x' \) and the equation solved to estimate \( \hat{q} \) as

\[
\hat{q} = \frac{x' - p}{1 - p}.
\]

Because false positives were accounted for, any value of \( p \) could be used. Stringent p-values would lead to an underestimate of the number of QTL because some real QTL would not be significant. However, a lenient p-value would not lead to overestimation of the number of QTL. In this analysis a range of p-values have been investigated but predominantly two values have been used, 0.1 and 0.5, to give an indication of the sensitivity of the results to the choice of \( p \). However, regardless of the p-value chosen, some QTL will have an effect too small to be significant and will not be counted.

The estimates of the numbers of real \( (\hat{r}) \) and false \( (\hat{f}) \) QTL were calculated as

\[
\hat{r} = \hat{q}n
\]

\[
\hat{f} = x'n - \hat{q}n
\]

where \( n \) is the total number of tests. The FDR was calculated as
\[ FDR = \frac{x' - \hat{q}}{x'}. \]

This is the same calculation as used by Mosig et al. (2001) except that this method does not use an iterative approach to estimate \( n_2 \) (true cases of their being no QTL). The number of heterozygous QTL per sire per trait was calculated as \((\hat{r}/6)/7\), because there were six sires and seven traits.

**Estimation of the proportion of heterozygous QTL, accounting for shadow QTL**

Positions that were at least 15cM apart were used to minimise the probability that a QTL caused a significant result in two or more adjacent marker intervals. However, this may still have occurred, causing some QTL to have been double counted and so overestimating the number of QTL. Significant, but false, QTL caused by a real QTL in an adjacent marker bracket will be referred to as shadow QTL. If there were no shadow QTL one would expect the significance tests in two adjacent marker brackets to be independent. To test whether any shadow QTL occurred, which would result in QTL being double counted, it was determined for all positions on a chromosome, bar the first, whether the QTL was significant at that position \((b)\) and the position before it \((a)\). Thus pairs of marker brackets in the same sire were classified into four possible categories, i.e., yes yes, yes no, no yes or no no, where yes meant a significant QTL. The observed frequencies of
these classes were calculated, as were the expected frequencies (assuming independence between adjacent marker brackets) as

\[
\begin{align*}
E(\text{yesyes}) &= \frac{a_y b_y}{n} \\
E(\text{yesno}) &= \frac{a_y b_n}{n} \\
E(\text{noyes}) &= \frac{a_n b_y}{n} \\
E(\text{nono}) &= \frac{a_n b_n}{n}
\end{align*}
\]

where \(a_y\) and \(b_y\) is the total number of observations significant at positions \(a\) and \(b\) respectively, \(a_n\) and \(b_n\) is the total number of observations not significant at position \(a\) and \(b\) respectively and \(n\) is the total number of observations. Using a \(\chi^2\) test the independence of the observed data was tested to determine whether in fact QTL had been double counted.

Once double counting was confirmed a new model for determining the proportion of heterozygous QTL was used, not only accounting for false positives but also for shadow QTL. Let \(q\) be the probability of a heterozygous QTL and \(p\) be the significance level. It was assumed that the QTL at positions \(a\) and \(b\) occurred independently of each other, and the probabilities of each of the heterozygous QTL status scenarios would be calculated as
\[
P(QQ) = q^2
\]
\[
P(Q0) = q(1 - q)
\]
\[
P(0Q) = q(1 - q)
\]
\[
P(00) = (1 - q)^2
\]

where \( Q \) is a heterozygous QTL and 0 is not. Where there were no heterozygous QTL, i.e., 00, a proportion \( p \) would still be significant. These false positives would not occur randomly. If there was a false positive at one position of the pair let \( s = P(\text{false } +ve \mid \text{false } +ve \text{ at other position}) \), then the probability of observing each significance scenario would be

\[
P(\text{yesyes}) = ps
\]
\[
P(\text{yesno}) = p - ps
\]
\[
P(\text{nryes}) = p - ps
\]
\[
P(\text{nono}) = 1 - 2p + ps.
\]

Where there is one heterozygous QTL, i.e. \( Q0 \) or \( 0Q \), it is likely to generate a significant result at the other position in the same way as one false positive generated another. This implies that an effect at one bracket will increase the probability of a false positive at the neighbouring bracket, whether the first effect is real or a false positive. Therefore for the heterozygous QTL status \( Q0 \) or \( 0Q \) it was assumed that

\[
P(\text{yesyes}) = s
\]
\[
P(\text{yesno}) = (1 - s).
\]

Therefore the total observed significance status probabilities were

\[
P(\text{nono}) = (1 - q)^2(1 - 2p + ps)
\]
\[
P(\text{nryes}) = 2q(1 - q)(1 - s) + 2(1 - q)^2(p - ps)
\]
\[
P(\text{yesyes}) = q^2 + 2q(1 - q)s + (1 - q)^2 ps
\]
where \( P(n\text{oyes}) \) included those in yes no class. Then

\[
P(data \mid q, s) = \prod P(ab)^{n_{ab}} \frac{(\sum n_{ab})!}{n_{\text{nono}}!n_{\text{n\text{oyes}}}!n_{\text{yesyes}}!}
\]

\[
\log L = \log P(data \mid q, s) = \sum (n_{ab} \ast \log P(ab) + \text{constant})
\]

The ML method found the estimates of \( \hat{q} \) and \( \hat{s} \) that maximised the logL subject to the constraints that \( 0 < \hat{q} < 1, 0 < \hat{s} < 1, 0 < P(ab) < 1 \) and \( \sum P(ab) = 1 \).

The effect of a heterozygous QTL on a trait is the difference between the effects of the two chromosomes and can be arbitrarily labelled as positive (+yes) or negative (-yes). When there are two QTL on adjacent segments the same chromosome can carry the positive allele for both QTL, here labelled as +yes+yes, or the positive allele at one QTL and the negative allele at the other QTL, here labelled as +yes-yes. If the QTL are in linkage equilibrium, these two possibilities are equally likely, so that

\[
P(+Q + Q) = \frac{q^2}{2}
\]
\[
P(+Q - Q) = \frac{q^2}{2}.
\]

However, it was assumed that a shadow QTL was always in the same direction as the neighbouring QTL so that all false positives led to +yes +yes, but could not lead to +yes –yes. The new observed significance status probabilities were
\[ P(\text{nono}) = (1-q)^2 (1-2p+ps) \]
\[ P(\text{noyes}) = 2q(1-q)(1-s) + 2(1-q)^2 (p-ps) \]
\[ P(\text{+yes} + \text{yes}) = \frac{q^2}{2} + 2q(1-q)s + (1-q)^2 ps \]
\[ P(\text{+yes} - \text{yes}) = \frac{q^2}{2}. \]

The ML method was solved for \( \hat{q} \) and \( \hat{s} \) in the same way as described above.

The two methods for estimating \( \hat{q} \) and \( \hat{s} \) (i.e., the method that distinguishes adjacent brackets with significant effects in opposite directions and the method that does not) represent opposite extremes in the models that could be used for the occurrence of real heterozygous QTL and shadow QTL. Therefore these estimates might be expected to bracket the correct estimates of \( q \) and \( s \).

**Calculation of the number of heterozygous sires for any given QTL**

This estimate of \( \hat{q} \), the proportion of marker brackets that contained a heterozygous QTL, did not determine the number of QTL that were segregating over the six sires for each trait. This is because it did not determine if QTL heterozygous in one sire are the same as those that were heterozygous in other sires. The purpose of the analysis described here was to do that.
Let \( n_i \) be the number of chromosome positions (i.e., marker brackets) where \( i \) sires had an apparently significant QTL at the desired significance threshold \( p \). Since there were six sires in total, observed values of \( n_0, n_1, n_2, n_3, n_4, n_5 \) and \( n_6 \) were calculated, and \( \sum n_i \) was equal to the number of chromosome positions analysed for all six sires. Let \( h_i \) be the probability that there were \( i \) sires that were really heterozygous at any given QTL, i.e., \( h_0, h_1, h_2, h_3, h_4, h_5 \) and \( h_6 \). Let \( x_i \) be the probability that \( i \) sires were significant, assuming that all sires that were really heterozygous were significant. So that

\[
x_i = \sum_{j=0}^{i} P(i \text{ sig. sires} \mid j \text{ real heterozygous sires})h_j
\]

\[
x_i = \sum_{j=0}^{i} P(i-j \text{ false +ve} \mid j \text{ real heterozygous sires})h_j
\]

\[
x_i = \sum_{j=0}^{i-1} \binom{6-j}{i-j} p^{i-j} (1-p)^{6-i} h_j.
\]

Then

\[
P(data \mid h's) = \prod_{i=0}^{6} x_i^{n_i} \frac{(\sum n_i)!}{n_0! n_1! n_2! n_3! n_4! n_5! n_6!}
\]

\[
LogL = \log P(data \mid h's) = \sum_{i=0}^{6} (n_i \log x_i) + \text{constant}.
\]

This ML method found the estimates of \( \hat{h}_0 \) to \( \hat{h}_6 \) that maximised the logL subject to the constraints that \( 0<h_i<1, \sum h_i = 1 \) and \( \hat{h}_4, \hat{h}_5 \) and \( \hat{h}_6 \) were set
to 0, because it was unlikely that so many sires were really heterozygous for any one QTL.

We were also able to determine the expected number of heterozygous sires, when at least one of \( n \) sires is heterozygous, under a neutral selection model (Crow and Kimura 1970) as

\[
\sum_{i=1}^{2n-1} \frac{1}{i} = n \frac{1}{2n-1}
\]

to determine if selection has occurred in this population.

RESULTS

If all null hypotheses were in fact true, that is there were no QTL that were really heterozygous, it would be expected that the p-values obtained from testing each marker bracket, in each sire, for each trait, would be uniformly distributed from zero to one. That is, the frequency of p-values would be the same in all intervals of 0.1, i.e., 287. However, the observed frequency, Table 1, was higher than expected (517) in the interval less than 0.1 and then progressively decreased to plateau off to a frequency of about 250 for intervals greater than 0.5. Therefore it can be concluded that there were in fact some QTL that were really heterozygous.
Estimation of the proportion of heterozygous QTL

Table 2 presents the results estimating the proportion of heterozygous QTL ($\hat{q}$), as well as the FDR, combining all sires and all traits, i.e., all 2870 data points. The table shows that as the significance threshold became more stringent, the proportion of observed significant QTL ($x'$) was reduced, as well as the estimate of the proportion of heterozygous QTL ($\hat{q}$). The real number of heterozygous QTL ($\hat{r}$), for each significance threshold, could be calculated using the formula $\hat{r} = \hat{q} n$, this enabled the calculation of the number of QTL segregating per sire per trait (Table 2). It is clear from Table 2 that the estimated number of real heterozygous QTL plateaus as the significance threshold becomes less stringent.

Estimating the proportion of heterozygous QTL, accounting for false positives and shadow QTL

The $\chi^2$ test for independence of significant QTL at adjacent chromosome positions was highly significant. Therefore it was concluded that some QTL had been double counted. Results for estimating the proportion of heterozygous QTL, accounting for false positives and shadow QTL, using a significance threshold of $p<0.1$ are presented in Table 3 and 4, where Table 4 distinguishes between two QTL of the same orientation and different orientation. From these tables it can be seen that the proportion of heterozygous QTL lay somewhere between 5 and 8% and the proportion of
shadow QTL, given a real QTL in the adjacent marker bracket, between 66 and 69%.

Table 3 and 4 also present the results for estimating the proportion of heterozygous QTL, accounting for false positives and shadow QTL, using a significance threshold of p<0.5. From these tables it can be seen that the proportion of heterozygous QTL lay somewhere between 8 and 27%, and the proportion of shadow QTL between 77 and 80%.

**Estimation of the number of heterozygous sires for any given QTL**

Results from the ML analysis estimating the number of heterozygous sires for any given QTL, using the observations of significant (p<0.5) QTL amongst the 49 marker brackets tested in all six sires, are presented in Table 5. It shows the number of QTL observed to be significant in *i* sires (*n_i*) and the estimated proportion of QTL really heterozygous for *i* sires (*h_i*). 16% of chromosome positions were not real heterozygous QTL in any of the six sires (*h_0*), and 84% of chromosome positions were real heterozygous QTL in only one of the six sires (*h_1*). These estimates changed very little when the constraint that *h_4*, *h_5* and *h_6* should be zero was removed, so that *h_0*=19%, *h_1*=80% and *h_6*=1%. 
Table 5 also shows the results using a more stringent significance threshold of $p<0.1$. Again most chromosome positions were real heterozygous QTL in only one of the six sires (55%), though a larger proportion of chromosome positions were not real heterozygous QTL in any of the six sires (36%), and 8% of positions were real heterozygous QTL in three sires. These estimates did not change at all when the constraint that $\hat{h}_4$, $\hat{h}_5$ and $\hat{h}_6$ should be zero was removed.

Under a neutral selection model the expected number of heterozygous sires, when at least one of 6 sires are heterozygous is 1.99.

**DISCUSSION**

This method calculated the number of heterozygous QTL using FDR methodology and could be applied to any data set. The methodology for calculating the FDR was the same as Mosig et al. (2001) except that our method did not have to use an iterative approach to estimate $n_2$ (true cases of there being no QTL). No matter what the method used to estimate the number of heterozygous QTL, there will always be some QTL too small to detect (BOVENHUIS and SCHROOTEN 2002). To minimise the number of false negatives a significance threshold of $p<0.5$ has been chosen to maximise the power of the experiment. A more stringent $p$ or equivalently a lower FDR would be used if we wanted to be confident that individual QTL were real.
However, that was not the purpose of this analysis; here the aim was to estimate the total number of real QTL controlling a trait. Therefore from Table 2, using $p<0.5$ and accounting for false positives but not shadow QTL, the proportion of marker brackets containing heterozygous QTL was 13%, which equates to approximately nine QTL segregating per sire per trait. This estimate of 9 might be an underestimate because the 89 marker brackets used did not cover the whole genome, only 21 of the 30 chromosomes were covered. Also, as stated previously, $p<0.5$ was chosen to maximise the power of the experiment and minimise false negatives. However, the power was still less than 1. Therefore our estimate of 9 could be underestimated. Though this number does agree well with the estimate of 10 heterozygous QTL per sire per trait estimated by Hayes and Goddard (2001) using a meta analysis of all the published data for dairy cattle.

This estimate, along with that of Hayes and Goddard (2001), did not correct for shadow QTL. It was possible that a QTL caused a significant test statistic in an adjacent marker bracket leading to a shadow QTL, and therefore some QTL may have been double counted. In this study midpoints 15cM apart were used to avoid double counting QTL, however, a significant $\chi^2$ test revealed that double counting had occurred. An alternative, but less likely, explanation is that QTL really do occur more commonly than expected in adjacent marker brackets. Therefore the proportion of shadows
was then estimated and the proportion of heterozygous QTL \( \hat{q} \) re-estimated, accounting for both false positives and shadow QTL (Tables 3 and 4). Using a significance threshold of \( p<0.1 \) the proportion of heterozygous QTL was found to be between 5 and 8\% (Table 3 and 4), whereas previously, not accounting for shadow QTL, it was 9\% (Table 2, \( p<0.1 \)). Using a significance threshold of \( p<0.5 \), so as to avoid missing QTL, the proportion of heterozygous QTL was found to be between 8 and 27\% (Table 3 and 4), when previously it was 13\% (Table 2, \( p<0.5 \)). The estimate of 27\% heterozygous QTL was inflated because, at a significance threshold of \( p<0.5 \), the number of observations that fell into the +yes-yes class was increased, compared with \( p<0.1 \). The assumption that observations that fell into the +yes-yes class must indicate two QTL being heterozygous resulted in the estimated number of QTL being inflated by false positives. However, using the \( p<0.1 \) significance threshold, there were very few of these +yes – yes observations because estimated effects at adjacent positions were positively correlated. Therefore the estimate of the proportion of heterozygous QTL \( \hat{q} \) was somewhere between 5 and 13\% (best estimate of 8\%), and the number of QTL per sire per trait between three and nine (best estimate five). A weakness of this analysis is that it does not account for multiple QTL within a marker bracket. Therefore, if there are indeed multiple QTL within a marker bracket, the number of QTL will have been underestimated. We have also assumed the presence of QTL in adjacent
brackets is independent. Clustering of QTL at distances greater than 15cM is possible but seems unlikely.

From the above analysis it was not known if the QTL that were segregating in one sire were also segregating in other sires. The ML estimation of the number of heterozygous sires for any given QTL (Table 5) indicates that most QTL were only heterozygous in one of the six sires, whether p<0.1 or p<0.5 was used.

Other papers have reported the average number of grandsires that appear to be heterozygous for significant QTL. Their estimates range from one in two, to one in six (Bennewitz et al. 2003; Bennewitz et al. 2004; Boichard 1997; Boichard et al. 2003; de Koning et al. 2001; Heyen et al. 1999; Zhang et al. 1998). These papers conducted stringent marker-trait tests for QTL, summing over all sires/grandsires, for daughter and granddaughter designs respectively. These tests are more likely to be significant if several sires/grandsires are heterozygous for the QTL. Therefore QTL that are heterozygous in only one of the sires/grandsires are generally not detected. This is evident in the papers mentioned above, in which QTL were rarely detected where less than two grandsires were heterozygous, particularly for QTL that were significant at the genome-wide and experiment-wide levels. Therefore, these papers have possibly overestimated the number of
heterozygous grandsires because they have not detected QTL that are heterozygous in only one grandsire. This bias has been avoided in the method presented here, because it has used sire-marker-trait tests, opposed to marker-trait tests, to determine the presence of a QTL and therefore a heterozygous sire. However, allowing for the bias, the estimates of one in two, to one in six grandsires being heterozygous for the detected QTL are not so different from the estimate of 1.0 to 1.25 (the average number of heterozygous sires where a QTL is detected) in six sires obtained here. This estimate is below that expected from a neutral selection model (1.99). Because the power to detect heterozygous QTL is less than 1 we will have underestimated the number of marker brackets with multiple heterozygous sires. However we will also have underestimated the number of marker brackets with only one heterozygous sire, and so our estimate of the average number of heterozygous sires (1.0 to 1.25) should not be badly biased. Therefore the simplest explanation is that selection has decreased the frequency of heterozygotes for QTL effecting milk production traits. This would be the expected consequence of natural selection acting against mutant alleles and causing lower frequency of mutant alleles than expected under the neutral model.

Assuming the estimate of one in six sires applies to all positions in the genome, not just those used in the analysis presented here, there were
between 18 (6*3) and 54 (6*9), best estimate of 30, QTL per trait segregating in the six sires. As the power of the experiment increases and the number of sires tested increases, the number of QTL detected increases, so these estimates should be regarded as minimum estimates. Hayes and Goddard (2001) adjusted the estimate of Lande (1981), to allow for unequal allelic effects, and the number of loci affecting a quantitative trait in an outbred population worked out to be 30, agreeing with the results presented here.

Since ASI is an index of 3 traits, it might be assumed that ASI should have 3 times as many QTL as components traits milk, fat and protein yields. However some QTL affect all 3 traits and some affect the components in a manner that leads to (almost) no effect on ASI. Therefore the number of QTL affecting ASI is less than 3 times the number affecting milk fat and protein yields. In addition, a QTL with small but detectable effect on one of the three component traits may have such a small effect on ASI that it is no longer detectable.
We would like to acknowledge all staff from the Molecular Genetics Department of Animal Genetics and Genomics, Department of Primary Industries Victoria, for their assistance in generating the large numbers of genotypes required as part of this work. This was completed as part of a joint venture between the Department of Primary Industries Victoria and AgResearch New Zealand. Pedigree and phenotypic information was supplied by the Australian Dairy Herd Improvement Scheme. The project was funded by the Victorian Governments Our Rural Landscape Initiative.
LITERATURE CITED

BEAVIS, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250-266 in *49th annual corn and sorghum industry research conference*, American Seed Trade Association, Washington, DC.


BLOTT, S., J. J. KIM, S. MOISIO, A. SCHMIDT-KUNTZEL, A. CORNET et al., 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.
BOICHARD, D., AND BISHOP, M.D., 1997 Detection of QTLs influencing milk production and mastitis resistance with a granddaughter design in Holstein cattle, pp. in 48th Annual Meeting of the EAAP, Commission on Animal Genetics, Vienna Austria.


BOVENHUIS, H., and C. SCHROOTEN, 2002 Quantitative Trait Loci for Milk Production Traits in Dairy Cattle, pp. in 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.


Academy of Sciences of the United States of America 87: 9888-9892.


Mosig, M. O., E. Lipkin, G. Khutoreskaya, E. Tchourzyna, M. Soller et al., 2001 A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of
selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. Genetics 157: 1683-1698.


ZHANG, Q., D. BOICHARD, I. HOESCHELE, C. ERNST, A. EGGEN et al., 1998

Table 1. Frequency of p-values from the data set, the expected frequency for each interval, if all null hypotheses were true, was 287.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-0.1</td>
<td>517</td>
</tr>
<tr>
<td>0.1-0.2</td>
<td>291</td>
</tr>
<tr>
<td>0.2-0.3</td>
<td>294</td>
</tr>
<tr>
<td>0.3-0.4</td>
<td>268</td>
</tr>
<tr>
<td>0.4-0.5</td>
<td>256</td>
</tr>
<tr>
<td>0.5-0.6</td>
<td>240</td>
</tr>
<tr>
<td>0.6-0.7</td>
<td>252</td>
</tr>
<tr>
<td>0.7-0.8</td>
<td>282</td>
</tr>
<tr>
<td>0.8-0.9</td>
<td>237</td>
</tr>
<tr>
<td>0.9-1.0</td>
<td>233</td>
</tr>
</tbody>
</table>
Table 2. Results for the calculation of the number of heterozygous QTL per sire per trait, as well as the FDR.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(p-value)</td>
<td>Sig</td>
<td>((\hat{r}))</td>
<td>((\hat{f}))</td>
<td>((\hat{q}))</td>
<td>/sire-trait</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.57</td>
<td>382</td>
<td>1244</td>
<td>0.77</td>
<td>0.13</td>
<td>9</td>
</tr>
<tr>
<td>0.2</td>
<td>0.28</td>
<td>293</td>
<td>516</td>
<td>0.64</td>
<td>0.10</td>
<td>7</td>
</tr>
<tr>
<td>0.1</td>
<td>0.18</td>
<td>256</td>
<td>261</td>
<td>0.51</td>
<td>0.09</td>
<td>6</td>
</tr>
<tr>
<td>0.05</td>
<td>0.12</td>
<td>210</td>
<td>133</td>
<td>0.39</td>
<td>0.07</td>
<td>5</td>
</tr>
<tr>
<td>0.01</td>
<td>0.04</td>
<td>96</td>
<td>28</td>
<td>0.22</td>
<td>0.03</td>
<td>2</td>
</tr>
<tr>
<td>0.005</td>
<td>0.03</td>
<td>71</td>
<td>14</td>
<td>0.16</td>
<td>0.02</td>
<td>2</td>
</tr>
<tr>
<td>0.001</td>
<td>0.01</td>
<td>28</td>
<td>3</td>
<td>0.09</td>
<td>0.01</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3. Observations and ML estimates of the proportion of heterozygous QTL ($\hat{q}$) and shadow QTL ($\hat{s}$), using data based on significance thresholds of $p<0.1$ and $p<0.5$.

<table>
<thead>
<tr>
<th>Class</th>
<th>$p&lt;0.1$</th>
<th>$p&lt;0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>no no</td>
<td>1673</td>
<td>721</td>
</tr>
<tr>
<td>no yes</td>
<td>195</td>
<td>439</td>
</tr>
<tr>
<td>yes yes</td>
<td>302</td>
<td>1010</td>
</tr>
</tbody>
</table>

Proportion real ($\hat{q}$)  0.06  0.09  
Proportion shadows ($\hat{s}$)  0.69  0.80
Table 4. Observations and ML estimates of the proportion of heterozygous QTL ($\hat{q}$) and shadow QTL ($\hat{s}$), distinguishing between 2 QTL in same and opposite orientation, using data based on a significance threshold of $p<0.1$ and $p<0.5$.

<table>
<thead>
<tr>
<th>Class</th>
<th>$p&lt;0.1$</th>
<th>$p&lt;0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>no no</td>
<td>1673</td>
<td>721</td>
</tr>
<tr>
<td>no yes</td>
<td>195</td>
<td>439</td>
</tr>
<tr>
<td>+yes +yes</td>
<td>263</td>
<td>805</td>
</tr>
<tr>
<td>+yes -yes</td>
<td>39</td>
<td>205</td>
</tr>
</tbody>
</table>

Proportion real ($\hat{q}$) 0.08 0.27

Proportion shadows ($\hat{s}$) 0.66 0.77

Log Likelihood -723.81 -1297.79
Table 5. Observations and ML estimates of the proportion of heterozygous QTL ($\hat{q}$) and shadow QTL ($\hat{s}$), distinguishing between 2 QTL in same and opposite orientation, using data based on significance thresholds of $p<0.1$ and $p<0.5$.

<table>
<thead>
<tr>
<th>Class</th>
<th>$p&lt;0.1$</th>
<th>$p&lt;0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>no no</td>
<td>1673</td>
<td>721</td>
</tr>
<tr>
<td>no yes</td>
<td>195</td>
<td>439</td>
</tr>
<tr>
<td>+yes +yes</td>
<td>263</td>
<td>805</td>
</tr>
<tr>
<td>+yes -yes</td>
<td>39</td>
<td>205</td>
</tr>
</tbody>
</table>

Proportion real ($\hat{q}$) 0.08 0.27

Proportion shadows ($\hat{s}$) 0.66 0.77
Table 6. Observations and ML estimates of the proportion of sires really heterozygous for any given QTL, using data based on significance thresholds of $p<0.1$ and $p<0.5$.

<table>
<thead>
<tr>
<th>No. sires (i)</th>
<th>Observed significant $(n_i)$</th>
<th>Proportion really het. $(\hat{h}_i)$</th>
<th>Observed significant $(n_i)$</th>
<th>Proportion really het. $(\hat{h}_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
<td>0.36</td>
<td>0</td>
<td>0.16</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>0.56</td>
<td>3</td>
<td>0.84</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.08</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>49</strong></td>
<td><strong>1</strong></td>
<td><strong>49</strong></td>
<td><strong>1</strong></td>
</tr>
</tbody>
</table>