Original article

Genes influencing milk production traits predominantly affect one of four biological pathways

Amanda Jane CHAMBERLAIN*, Helen Clare MCPARTLAN,
Michael Edward GODDARD

Animal Genetics and Genomics Platform, Department of Primary Industries,
Victoria, Australia

Abstract – In this study we introduce a method that accounts for false positive and false negative results in attempting to estimate the true proportion of quantitative trait loci that affect two different traits. This method was applied to data from a genome scan that was used to detect QTL for three independent milk production traits, Australian Selection Index (ASI), protein percentage (P%) and fat percentage corrected for protein percentage (F% – P%). These four different scenarios are attributed to four biological pathways: QTL that (1) increase or decrease total mammary gland production (affecting ASI only); (2) increase or decrease lactose synthesis resulting in the volume of milk being changed but without a change in protein or fat yield (affecting P% only); (3) increase or decrease protein synthesis while milk volume remains relatively constant (affecting ASI and P% in the same direction); (4) increase or decrease fat synthesis while the volume of milk remains relatively constant (affecting F% – P% only). The results indicate that of the positions that detected a gene, most affected one trait and not the others, though a small proportion (2.8%) affected ASI and P% in the same direction.

bivariate analysis / independent traits / pleiotropy / genome scan / false discovery rate

1. INTRODUCTION

As a result of complex biochemical, developmental and regulatory pathways, a polymorphism in a single gene will almost always influence multiple traits, a phenomenon known as pleiotropy [6, 8]. In classical quantitative genetics, pleiotropy is recognised by genetic correlations. The existence of a genetic correlation between two traits implies that some genes must affect both

---

traits, or that two genes affecting two different traits are in linkage disequilibrium. However a low genetic correlation might mean that few genes affect both traits, or that many genes affect both traits, but sometimes in the same direction and other times in the opposite direction. In the case of many genes affecting both traits an increase in one trait would not necessarily mean an increase or decrease in the other. Knowledge of the pattern of pleiotropy would increase our understanding of the biology underlying quantitative genetic variation and would help us to identify the genes causing variation in quantitative traits (QTL).

Experiments mapping QTL should be able to describe the pattern of pleiotropy across the traits studied, but they rarely report results in this way. An exception is Lipkin *et al.* [5] who reported QTL affecting milk yield (M), protein yield (P) and protein percentage (P%) in dairy cattle. They found that of the QTL affecting at least one trait, 11% were significant for one trait only, 25% for two traits and 64% for all three traits. However, when QTL are tested, some will be significant by chance alone (false positives) and some that are real will not be significant (false negatives). Since protein percentage (P%) is equal to protein yield (P) divided by milk yield (M) a gene must affect at least two of these traits. However the occurrence of false negatives means that QTL can have a significant effect on only one of the three traits.

In this study we introduce a method that accounts for false positive and false negative results in attempting to estimate the true proportion of QTL that affect two different traits. This method was applied to data from a genome scan that was used to detect QTL for three independent milk production traits. The results show that most QTL affect only one of four biological pathways involved in milk production.

2. MATERIALS AND METHODS

The raw data for the analysis reported in this paper comes from a QTL mapping experiment. QTL express [7] was used to perform linkage analysis on genotype data from a selective genotyping experiment consisting of six sires, each with approximately 100 daughters selected for high and 100 daughters selected for low Australian Selection Index (ASI). ASI is an economic index of milk, fat and protein yields, where ASI = (3.8*protein) + (0.9*fat) − (0.048*milk). Protein percentage, fat percentage and ASI phenotypes were provided by the Australian Dairy Herd Improvement Scheme as deregressed Australian Breeding Values (ABV). Fat percentage corrected for protein percentage was calculated as (F% – P%) = F% – 1.4299P% based on a regression

analysis performed for fat percentage phenotypes on protein percentage phenotypes. ASI, protein percentage (P%) and F% – P% were used in the linkage analysis. The correlation between deregressed ABV for ASI and P% was found to be 0.07, 0 for ASI and F% – P% and P% and P% – F%, and this greatly simplifies the analysis described below.

Linkage analyses were fixed at marker midpoints throughout the genome. Only positions greater than 15 cM apart were chosen for this analysis, with data extracted from the QTL express output. Each record consisted of the results from one chromosome position in one sire family and contained the p-values from the t-test for a QTL at that position affecting ASI, P% and F% – P%. Therefore the final data set consisted of three signed p-values (one for each trait) for a total of 89 chromosomal positions, at least 15 cM apart, for between one and six sires (an average of 4.6 sires per position), giving a total of 410 QTL tests.

The aim of the analysis was to estimate, for each pair of traits, the proportion of QTL that affected neither trait, one trait but not the other, both traits in the same direction or both traits in opposite directions. Real heterozygous QTL could be allocated into five different categories based on the effects they have on two independent traits. A QTL could have an effect on both traits; here these fell into one of two categories, 1 or 3. Since QTL effects were arbitrarily estimated as either positive or negative those that fell into category 1 could have either a positive or negative effect on both traits, *i.e.*, they could affect both traits in the same orientation. Those that fell into category 3 have an effect on both traits, but, the effects were in the opposite orientation, *e.g.*, positive for trait 1 while negative for trait 2. Alternatively, a QTL could have an effect on one trait (positive or negative) while having no effect on the other trait, here these fell into categories 2 and 4 for the two different traits respectively. The majority of chromosome positions have no effect on either trait, and here these fell into category 5. The probabilities of these categories, r_1, r_2, r_3, r_4 and r_5 are the probability of a real heterozygous QTL having those effects on the two traits.

However, a significant effect at a given position can be a false positive and this needs to be taken into account when estimating the number of real QTL. An observed QTL could have a significant effect, where p is less than some threshold, on both traits, categories, 1 or 3. Those that fell into category 1 had a significant effect that was either positive or negative for both traits, *i.e.*, they effected both traits in the same orientation. Those that fell into category 3 had a significant effect on both traits, however, the effects were in the opposite orientation. Alternatively an observed QTL could have had a significant effect on

one trait (positive or negative) while not being significant for the other trait, categories 2 and 4. The majority of loci were not significant for either trait, category 5. The probabilities of these categories, x_1, x_2, x_3, x_4 and x_5, represented in Figure 1, are the probabilities of observing a significant QTL having those effects on the two traits.

Real QTL effects were combined with what were observed, resulting in Figure 2. It was important to use traits that were independent of one another, so that a false positive for one trait did not change the probability of a false positive for the other trait. Then, if it assumed that real QTL are always significant, it was possible to model the probabilities of each observed category in terms of r and p as shown in Figure 2. From Figure 2, equations for the probabilities of observing categories 1, 2, 3, 4 and 5 were derived as

$$x_1 = r_1 + \frac{p}{2}r_2 + \frac{p}{2}r_4 + \frac{p^2}{2}r_5$$

$$x_2 = (1 - p)r_2 + p(1 - p)r_5$$

$$x_3 = r_3 + \frac{p}{2}r_2 + \frac{p}{2}r_4 + \frac{p^2}{2}r_5$$

$$x_4 = (1 - p)r_4 + p(1 - p)r_5$$

$$x_5 = (1 - p)^2 r_5.$$

Using the observed (x) values, equations were solved for $\hat{r}_5$, then $\hat{r}_4$, $\hat{r}_2$, $\hat{r}_3$ and $\hat{r}_1$.

In matrix notation $\hat{r} = Tx$,

$$
\begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \hat{r}_3 \\ \hat{r}_4 \\ \hat{r}_5 \end{bmatrix}
=
\begin{bmatrix}
1 & -\dfrac{p}{2(1-p)} & 0 & -\dfrac{p}{2(1-p)} & \dfrac{p^2}{2(1-p)^2} \\
0 & \dfrac{p}{(1-p)} & 0 & 0 & \dfrac{p^2}{(1-p)^2} \\
0 & -\dfrac{p}{2(1-p)} & 1 & -\dfrac{p}{2(1-p)} & \dfrac{p^2}{2(1-p)^2} \\
0 & 0 & 0 & \dfrac{1}{(1-p)} & -\dfrac{p}{(1-p)^2} \\
0 & 0 & 0 & 0 & \dfrac{1}{(1-p)^2}
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}
$$

Trait 1

	p<threshold (+ve)	p>threshold	p<threshold (−ve)
Trait 2 — p<threshold (+ve)	x_1	x_2	x_3
Trait 2 — p>threshold	x_4	x_5	x_4
Trait 2 — p<threshold (−ve)	x_3	x_2	x_1

Figure 1. Observed significant QTL effecting two traits.

Trait 1

		p<threshold (+ve)		p>threshold	p<threshold (−ve)	
		real	false	false	real	false
Trait 2 — p<threshold (+ve)	real	r_1	$\frac{p}{2}r_2$	$(1-p)r_2$	r_3	$\frac{p}{2}r_2$
			x_1	x_2	x_3	
	false	$\frac{p}{2}r_4$	$\frac{p^2}{2}r_5$	$p(1-p)r_5$	$\frac{p}{2}r_4$	$\frac{p^2}{2}r_5$
Trait 2 — p>threshold	false	$(1-p)r_4$	x_4 $p(1-p)r_5$	x_5 $(1-p)^2r_5$		

Figure 2. Combined real heterozygous and observed significant QTL effecting two traits.

The variances of the $\hat{r}$'s were represented by the diagonal elements of the matrix $V(\hat{r}) = TV(x)T'$, where $V(x)$ is

$$\begin{bmatrix} V(x_1) \\ V(x_2) \\ V(x_3) \\ V(x_4) \\ V(x_5) \end{bmatrix} = \begin{bmatrix} \dfrac{x_1(1-x_1)}{n} & -\dfrac{x_1 x_2}{n} & -\dfrac{x_1 x_3}{n} & -\dfrac{x_1 x_4}{n} & -\dfrac{x_1 x_5}{n} \\ -\dfrac{x_2 x_1}{n} & \dfrac{x_2(1-x_2)}{n} & -\dfrac{x_2 x_3}{n} & -\dfrac{x_2 x_4}{n} & -\dfrac{x_2 x_5}{n} \\ -\dfrac{x_3 x_1}{n} & -\dfrac{x_3 x_2}{n} & \dfrac{x_3(1-x_3)}{n} & -\dfrac{x_3 x_4}{n} & -\dfrac{x_3 x_5}{n} \\ -\dfrac{x_4 x_1}{n} & -\dfrac{x_4 x_2}{n} & -\dfrac{x_4 x_3}{n} & \dfrac{x_4(1-x_4)}{n} & -\dfrac{x_4 x_5}{n} \\ -\dfrac{x_5 x_1}{n} & -\dfrac{x_5 x_2}{n} & -\dfrac{x_5 x_3}{n} & -\dfrac{x_5 x_4}{n} & \dfrac{x_5(1-x_5)}{n} \end{bmatrix}$$

and the standard errors of the $\hat{r}$'s were calculated as $se(\hat{r}) = \sqrt{V(\hat{r})}$.

Equating observed proportions to expected proportions provides maximum likelihood estimates of the parameters in certain simple cases like these.

3. RESULTS

Because the method used here corrects for false positives, it is possible to use any significance level (p). We performed the calculations using both p < 0.1 and p < 0.5. The method assumes that all real QTL will be significant, so p < 0.1 may result in more false negatives than p < 0.5.

r_5 is the proportion of chromosome positions that had no effect on either trait, r_2 is the proportion of chromosome positions that really effected trait 2 and not trait 1, r_4 is the proportion of chromosome positions that really effected trait 1 and not trait 2, r_1 is the proportion of chromosome positions that really effected both traits in the same orientation and r_3 is the proportion of chromosome positions that really effected both traits in the opposite orientation. The results from three independent traits, ASI, P% and F% − P% using p < 0.1, are presented in Tables I, II and III. The positive/positive cell corresponds to the category r_1 and so also includes negative/negative QTL. The negative/positive cell corresponds to category r_3 and so also includes positive/negative QTL. The 0/positive cell corresponds to category r_2 and also includes 0/negative QTL. The positive/0 cell corresponds to category r_4 and also includes negative/0 QTL. These results indicate that of the chromosome positions that were really heterozygous QTL, most had an effect on one trait and not the others (categories r_2 and r_4). Negative estimates were due to sampling error given that only 410 positions were used.

Table I. Estimates of r_1, r_2, r_3, r_4 and r_5 and their standard errors for Australian Selection Index (ASI) and protein percentage (P%), using significance thresholds of $p < 0.1$ (denoted by **) and $p < 0.5$ (denoted *).

		ASI		
		+ve	0	−ve
P%	+ve	2.8% ± 1.0%**	4.5% ± 1.9%**	−0.8% ± 0.3%**
		17.1% ± 3.1%*	−3.9% ± 6.5%*	−1.5% ± 2.4%*
	0	8.3% ± 2.2%**	85.2% ± 2.8%**	
		−2.4% ± 6.6%*	90.7% ± 8.3%*	

** $p < 0.1$, * $p < 0.5$.

Table II. Estimates of r_1, r_2, r_3, r_4 and r_5 and their standard errors for Australian Selection Index (ASI) and fat percentage corrected for protein percentage (F% − P%), using significance thresholds of $p < 0.1$ (denoted by **) and $p < 0.5$ (denoted *).

		ASI		
		+ve	0	−ve
F% − P%	+ve	−0.7% ± 0.4%**	6.0% ± 2.0%**	2.0% ± 0.9%**
		−1.3% ± 2.8%*	15.6% ± 6.5%*	0.4% ± 2.9%*
	0	9.0% ± 2.2%**	83.7% ± 2.9%**	
		14.6% ± 6.4%*	71.2% ± 7.6%*	

** $p < 0.1$, * $p < 0.5$.

Table III. Estimates of r_1, r_2, r_3, r_4 and r_5 and their standard errors for protein percentage (P%) and fat percentage corrected for protein percentage (F% − P%), using significance thresholds of $p < 0.1$ (denoted by **) and $p < 0.5$ (denoted *).

		P%		
		+ve	0	−ve
F% − P%	+ve	0.5% ± 0.6%**	5.1% ± 2.0%**	1.0% ± 0.7%**
		1.2% ± 2.9%*	14.6% ± 6.4%*	−4.2% ± 2.7%*
	0	5.9% ± 2.0%**	87.6% ± 2.8%**	
		17.1% ± 6.5%*	71.2% ± 7.6%*	

** $p < 0.1$, * $p < 0.5$.

Repeated analysis using a more relaxed significance threshold of $p < 0.5$, so as to minimise false negatives, for the three independent traits, ASI, P% and F% − P%, are also presented in Tables I, II and III. The results still show that most QTL really had an effect on one trait and not the others, except for those QTL that had an effect on ASI or P%. If the QTL affected ASI or P% then the QTL would most likely affect the other trait in the same direction (category r_1).

The apparent discrepancy between these results, and that using $p < 0.1$, could be explained if genes that had a large effect on ASI or P% tended to have a smaller effect on the other trait which is detected at $p < 0.5$, but not $p < 0.1$.

4. DISCUSSION

When three traits were considered, ASI, P% and F% – P%, it was found that QTL predominantly only affected one of the three, except QTL affecting ASI that also had a small effect on P% and *vice versa*. We used uncorrelated traits here so that a false positive for one trait does not increase the probability of a false positive for the other trait. Correlated traits would most definitely have found genes that affect more than one trait, since that is the nature of being correlated. The power of the experiment was maximised by reducing the stringency of the test (*i.e.*, to $p < 0.5$). However, the power was still likely to be less than 1 and so inevitably some QTL where missed. This may mean we underestimated the number of real QTL, however there is no reason to believe that the pattern of the missed QTL would be any different to those detected. This is, however, an area that could be investigated further.

Effects on each of these three traits correspond approximately to simple biological interpretations. QTL that affect only ASI, not P% or F% – P%, would result from increased or decreased mammary gland production of milk with no change in milk composition. In this case protein and fat yield would be increased in the same proportions and so the composition of the milk would remain relatively unchanged. QTL that affect only P% correspond to increased or decreased lactose synthesis within the mammary gland resulting in the volume of milk being changed in the same direction, but without a change in protein or fat yield. Protein synthesis would remain constant causing a change in the percentage of protein in the milk and also a small change in ASI. QTL that affect both ASI and P% in the same direction result from an increase or decrease in protein synthesis while milk volume remains relatively constant. QTL that affect only F% – P% correspond to increased or decreased fat synthesis within the mammary gland, again while the volume of milk remains relatively constant. The two trait analysis using $p < 0.1$ revealed that the QTL appear to only affect one of these three traits. When the significance threshold was relaxed to $p < 0.5$, QTL that had an effect on ASI also had an effect on P% in the same orientation. This could be explained if QTL that had a large effect on one of these traits did have a small effect on the other trait. For instance, genes affecting lactose synthesis do have a large effect on P% and a smaller effect on ASI, as pointed out already. Alternatively, genes that increase only

Table IV. The results of sire-marker-trait tests that were significant for two traits, protein yield (P) and protein % (P%) from Lipkin *et al.* [5].

		P		
		+ve	**0**	**−ve**
P%	**+ve**	1.0%	4.6%	0.5%
	0	2.1%	91.8%	

Table V. The results of sire-marker-trait tests that were significant for two traits, milk yield (P) and protein percentage (P%) from Lipkin *et al.* [5].

		M		
		+ve	**0**	**−ve**
P%	**+ve**	0.5%	4.6%	2.1%
	0	3.3%	89.5%	

protein synthesis would increase ASI and P% in the same proportions. However, genes with a large effect on both of these traits appear to be rare. Only the large effects were detected in the more stringent test using $p < 0.1$.

These results support the published literature, where QTL effecting protein synthesis have been rarely reported. According to a literature review by Chamberlain [2] 0.94% of reported QTL affected protein yield (P) and protein % (P%). It has been shown that some genes affect milk production traits, these also support our results. DGAT1 [4] predominantly affects fat synthesis, GHR [1] predominantly affects milk volume as does ABCG2 [3].

The study of Lipkin *et al.* [5] used three traits, milk yield (M), protein yield (P) and protein percentage (P%). They conducted 2844 sire-marker-trait tests. The results, presented in Tables IV and V, show that most sire-marker-trait tests appeared to affect protein yield, protein percentage or milk yield. However, Lipkin *et al.* used a false discovery rate of 0.05, which means that only the largest QTL effects would have been detected. The more lenient thresholds used here of $p < 0.5$ and $p < 0.1$ are equivalent to FDR of 0.77 and 0.51 respectively, so more QTL with smaller effect would have been detected, and so QTL with smaller effects on more than one trait were also more likely to be detected. Of the small number of sire-marker-trait tests that were significant for both P and P% (Tab. IV) most affected both traits in the same orientation. This would occur where protein synthesis is increased or decreased, while the volume of milk remained constant. Of the small number of sire-marker-trait tests that were significant for both M and P% (Tab. V) most affected the two traits in the opposite orientation. This would occur where lactose synthesis is increased or decreased resulting in an increase in milk volume, while the

yield of protein would remain unchanged. These results were confirmed in the marker-trait tests, which summed across all sires, where most marker-trait tests were significant for two or more traits. It appeared that the relationship between the three traits was that a QTL affecting one trait must affect one of the other two traits, but not necessarily all three, and based on the results of the sire-marker-trait tests, not necessarily in all sires. As the power of the scan was less than 1, QTL found to affect only one or two traits could have been due to sampling error and in reality affect all three traits. However, QTL affecting all three traits is inevitable considering that protein percentage is actually protein yield/milk yield. This is a case of using correlated traits.

The results of Lipkin *et al.* [5] support the conclusion that QTL affecting lactose synthesis exist. These QTL have an effect on ASI or P% and also have a small effect on the other trait. However, QTL with a large affect on ASI and P%, resulting from an increase or decrease in protein synthesis, are rare.

In this analysis it was fortuitous that ASI and P% were uncorrelated. In general uncorrelated traits can be created by correcting one trait for the other as was done here for fat % corrected for protein %. Therefore the approach used here could be applied to correlated traits.

ACKNOWLEDGEMENTS

REFERENCES

[1] 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., 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 (2003) 253–266.

[2] Chamberlain A., Mapping quantitative trait loci for milk production traits in dairy cattle and using them in marker assisted selection, Institute of Land and Food Resources, University of Melbourne, Melbourne, 2004, pp. 218.

[3] Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., Everts-van der Wind A., Lee J.H., Drackley J.K., Band M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I., Ron M., Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle, Genome Res. 15 (2005) 936–944.

[4] Grisart B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M., Snell R., Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition, Genome Res. 12 (2002) 222–231.

[5] Lipkin E., Grosman G., Tchourzyna E., Friedmann A., Soller M., QTL affecting milk protein percent also affect milk yield and protein yield, in: Proceedings of 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, France.

[6] Lynch M., Walsh B., Genetics and Analysis of Quantitative Traits, Sinauer Associates, Inc., Sunderland, USA, 1998.

[7] Seaton G., Haley C.S., Knott S.A., Kearsey M., Visscher P.M., QTL Express: mapping quantitative trait loci in simple and complex pedigrees, Bioinformatics 18 (2002) 339–340.

[8] Wright S., Evolution and the genetics of populations, University of Chicago Press, Chicago, USA, 1968.