Statistical modelling of main and epistatic gene effects on milk production traits in dairy cattle

Joanna Szyda[1] and Jolanta Komisarek[2]

[1] *Department of Animal Genetics, Wrocław Agricultural University, Kożuchowska 7, 51-631 Wrocław, Poland*
[2] *Department of Cattle Breeding and Milk Production, August Cieszkowski Agricultural University of Poznań, Wojska Polskiego 71 A, 60-625 Poznań, Poland*

The study was partly financed by the Polish State Committee for Scientific Research, grant no. 2 P06D 017 26

INTRODUCTION

In the last decade, numerous QTL (quantitative trait loci) for milk production traits have been identified in various dairy cattle populations (Khatkar et al., 2004). Most of them were found on bovine chromosome 1, 3, 6, 14 and 20 (Ashwell et al., 2004; Plante et al., 2001). The less frequently reported QTL are located e.g. on chromosome 4 (Lindersson et al., 1998) and 23 (Ashwell et al., 1997). Although genes underlying these QTL variation mostly remain unknown, a number of potential candidate genes have been selected. The genes with a possible effect on milk traits in cattle can be involved in the different physiological pathways, such as fat synthesis (DGAT1) (Farese et al., 2000), fat secretion from the mammary epithelial tissue (BTN1A1) (Ogg et al., 2004), or whole body energy balance regulation (LEP and LEPR) (Houseknecht et al., 1998).

The aim of this study was to investigate effect and potential interactions of the bytyrophilin gene, the diacylglycerol acyltransferase 1 gene, the leptin gene and the leptin receptor, whose segregation was assessed by nine single nucleotide polymorphisns (SNP).

MATERIAL

The animal material comprises 252 Black-and-White bulls from the active Polish dairy population, born between 1990 and 1998. The number of daughters per bull was high with the average varying among bull's birth years between 81 and 4953 for the 1st lactation, as well as between 64 and 4033 for the 2nd lactation. For each bull the following functional single nucleotide polymorphisms were assessed: F16Y, P35Q and K468R in the bytyrophilin (*BTN1A1*) gene located on bovine chromosome 23, K232A in the diacylglycerol acyltransferase 1 (*DGAT1*) gene on chromosome 14, T945M in the leptin receptor (*LEPR*) gene on BTA3, and Y7F, R25C, A80V as well as promoter C/T substitution at position -963 in the leptin (*LEP*) gene on chromosome 4. For F16Y and Y7F only two genotypes (common homozygous and heterozygous) were observed whereas for the other polymorphisms all three genotypes were identified. Details on polymorphisms and their genotype frequencies are presented in Table 1.

Breeding values (EBV) estimated on a national basis using a random regression test day model and daughter yield deviations (DYD) derived following Liu et al. (2004) from the above model were considered as measures of bull's milk, fat and protein yield merits at 1st and 2nd lactations. Simple descriptive statistics for the trait measures are presented in Table 2.

Table 1 Single nucleotide polymorphisms considered in this study.

Code	Gene	Location	Genotypes and genotype frequencies		
T945M	leptin receptor (LEPR)	BTA03	TT=0.01	CT=0.15	CC=0.84
Y7F	leptin gene (LEP)	BTA04	AA=0.98	AT=0.02	
R25C			TT=0.16	CT=0.57	CC=0.27
A80V			CC=0.47	CT=0.45	TT=0.08
C(-963)T			CC=0.28	CT=0.57	TT=0.15
K232A	diacylglycerol acyltransferase 1 gene (DGAT1)	BTA14	AA/AA=0.456	AA/GC=0.267	GC/GC=0.277
F16Y	bytyrophilin gene (BTN1A1)	BTA23	TT=0.92	AT=0.08	
P35Q			AA=0.37	AC=0.47	CC=0.16
K468R			AA=0.79	AG=0.20	GG=0.01

Table 2 Summary of trait measures considered in this study.

Trait		Measure	
		DYD	EBV
Milk yield 1st parity	$\bar{x}$	298.5 kg	299.7 kg
	$\hat{\sigma}$	323.9 kg	308.8 kg
Milk yield 2nd parity	$\bar{x}$	353.3 kg	355.3 kg
	$\hat{\sigma}$	395.2 kg	371.0 kg
Protein yield 1st parity	$\bar{x}$	8.3 kg	8.3 kg
	$\hat{\sigma}$	8.5 kg	8.0 kg
Protein yield 2nd parity	$\bar{x}$	8.8 kg	8.9 kg
	$\hat{\sigma}$	11.8 kg	10.9 kg
Fat yield 1st parity	$\bar{x}$	8.4 kg	8.3 kg
	$\hat{\sigma}$	13.0 kg	12.2 kg
Fat yield 2nd parity	$\bar{x}$	9.5 kg	9.4 kg
	$\hat{\sigma}$	17.3 kg	16.1 kg

METHODS

Polymorphism assessment

DNA for molecular analyses was extracted from peripheral blood using the standard phenol method. Genotypes were determined using PCR-RFLP. Primers for the PCR (Table 3) were established on the basis of the gene sequences available in the GenBank data base (accession numbers: BTN1A1-Z93323, DGAT1-AY065621, LEP-U50365, LEPR-AJ580801), with the use of PRIMER3 software (http://www.genome.wi.mit.edu/cgibin/primer/primer3_www.cgi).

The PCR reaction mixture contained a total of 20–50 ng genomic DNA, 0.5 units of Taq polymerase (Fermentas, Vilnius, Lithuania), 1x PCR buffer with (NH4)2SO4, 2 mM MgCl2, 5 % DMSO, 1 μM of each primer (IBB PAS, Warsaw, Poland), and 200 μM of each dNTP (Fermentas, Vilnius, Lithuania). Thermal cycling conditions included an initial denaturation at 94 ^{0}C for 5 min, followed by 30 cycles of 94 ^{0}C for 30 s, annealing temperature (Table 3) for 30 s, and 72 ^{0}C for 40 s, followed by the final extension at 72 ^{0}C for 5 min. The amplified fragments were digested overnight with 5 units of respective (Table 3) restriction

endonucleases (Fermentas, Vilnius, Lithuania), and next subjected to electrophoretic separation in 2.5 % ethidium bromide stained agarose gel (BASICA LE GQT, Prona).

Table 3 Selected PCR-RFLP conditions for the analysed polymorphisms.

SNP	Primers (5'-3')	Annealing temp. (°C)	PCR product size	Restri-ction enzyme	Digestion product size
T945M	F: GCAACTACAGATGCTCTACTTTTGT* R: CAGGGAAATTTCCCTCAAGTTTCAA	56	400 bp	*Taq*I	T- 400 bp, C- 375, 25 bp
Y7F	F: CTGCGTGGTCTACAGCACACCTC R: AGGGCCAAAGCCACAGGAT*TCGA	57,5	310 bp	*Bsp119*I	A- 310 bp, T- 288, 22 bp
R25C	F: CCAGGGAGTGCCTTTCATTA R: GGTGTCATCCTGGACCTTC*C	56,5	305 bp	*Kpn2*I	T- 305 bp, C- 283, 22 bp
A80V	F: CAAGCAGGAAATAGGGAGTCATGG R: CTGGTGAGGATCTGTTGGTAGG*TC	56	424 bp	*Eco91*I	C- 424 bp, T- 398, 26 bp
C(-963)T	F: GTGATCAGAAAACACATACCATTTTATAAT R: GCCTGGTTGTTTTGCTTTTAATAATTATCTT*	55	295 bp	*Dra*I	C- 295 bp, T- 268, 27 bp
K232A	F: TGCCGCTTGCTCGTAGCTTTGGCC* R: ACCTGGAGCTGGGTGAGGAACAGC	58,5	378 bp	*Bgl*I	AA- 96, 282 bp, GC- 28, 96, 254 bp
F16Y	F: ATTGACTAACCTTAGGGTGGTAGGT R: TTGGGCAGCTGGAGGAGAATT*	57	320 bp	*Dra*I	T- 320bp, A- 298 and 22 bp
P35Q	F: TGGTAGGTCAGGAAGCCATC R: GTATTCAGCCATCTCCTCGC	58	574 bp	*Bcn*I	A- 74 and 500bp, C- 74, 96 and 404 bp
K468R	F: TGGAGCTCTATGGAAATGGG R: ACCCTTTGGGTTTTCTGCTT	56	780 bp	*BsuR*I	A- 10, 13, 83, 141, 162, 371 bp, G- 10, 13, 33, 83, 141, 162, 338 bp

* an intentional mismatch incorporating the restriction site to a sequence

Statistical modelling

Models: A series of mixed models were applied to the data with the general structure given by:

$$\mathbf{y} = \mathbf{X}_{\beta}\boldsymbol{\beta} + \mathbf{X}_{q}\mathbf{q} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e} \ ,$$

where **y** is a trait measure (DYD or EBV), **β** is a vector of fixed nongenetic effects comprising a general mean and a bull's birth year, **q** is a vector of fixed SNP effects, **α** is a vector of random polygenic effects assuming $\boldsymbol{\alpha} \sim N\left(0, \mathbf{A}\sigma_{\alpha}^{2}\right)$ with **A** representing additive relationships among individuals and σ_{α}^{2} being a component of the total additive genetic variance attributed to polygenes; **e** is a vector of random errors assuming $\mathbf{e} \sim \mathrm{N}\left(0, \mathbf{R}\sigma_{e}^{2}\right)$ with **R** being a diagonal matrix with the reciprocal of the number of daughters used for the calculation of DYD or EBV and σ_{e}^{2} denoting the error variance, $\mathbf{X}_{\beta}$, $\mathbf{X}_{q}$ and **Z** are corresponding design matrices. Both variance components were assumed as known (i.e. not estimated) with σ_{α}^{2}=0.3 and σ_{e}^{2}=0.6. Differences between applied models comprised **q**. The I-th SNP in **q** is modelled through its additive (a_i) and dominance (d_i) effects, additionally an additive-by-additive epistasis between i-th and j-th SNP (δ_{ij}) is considered in multiple SNPs models. Coding of the corresponding elements of $\mathbf{X}_{q}$ follows the F_{∞} metric model as defined by Kao and Zeng (2002), where:

$$x_{ai} = \begin{cases} 1 & \text{for a homozygous SNP (say QQ)} \\ 0 & \text{for a heterozygous SNP (say Qq)} \\ -1 & \text{for a homozygous SNP (say qq)} \end{cases}, \quad x_{di} = \begin{cases} 0 & \text{for QQ and qq} \\ 1 & \text{Qq} \end{cases}, \quad \text{and } x_{\delta ij} = x_{ai}x_{aj}.$$

In the analysis 46 models with the following parameterisation of **q** were considered:

a) No SNP model - $\mathbf{y} = \mathbf{X}_{\beta}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}$.

b) Nine single SNP models - $\mathbf{q}^T = \begin{bmatrix} a_i & d_i \end{bmatrix}$ for i-th SNP.

c) A model with all SNPs - $\mathbf{q}^T = \left[\sum_{i=1}^{9} \begin{pmatrix} a_i & d_i \end{pmatrix} \right]$.

d) Thirty-nine models with all SNPs and a single epistasis - $\mathbf{q}^T = \left[\sum_{i=1}^{9} \begin{pmatrix} a_i & d_i \end{pmatrix} \delta_{ii'} \right]$.

e) A model with all SNPs and all possible epistatic effects - $\mathbf{q}^T = \left[\sum_{i=1}^{9} \begin{pmatrix} a_i & d_i \end{pmatrix} \sum_{i=1}^{9} \sum_{j=i+1}^{9} \delta_{ij} \right]$.

Model selection: two criteria were considered for model selection:

a) The standard Bayesian Information Criterion (Schwarz, 1978): $\mathrm{BIC} = \ln L - \frac{1}{2} k \ln n$, where L represents model likelihood, k is the number of fitted parameters and n – the number of observations.

b) A version of BIC (mBIC) modified by Baierl et al. (2006) for the purpose of a genome scan (i.e. multiple linked markers), enabling differentiation between marginal (additive and dominance) marker effects and interaction (epistatic) effects between markers: $\mathrm{mBIC} = \ln L - (p+q)\ln n - 2p\ln\left(\frac{m}{1.1} - 1\right) - 2q\ln\left(\frac{m(m-1)}{1.1} - 1\right)$, where p and q are the number of fitted main and epistatic effects respectively and m is the number of markers.

RESULTS

Since the analysis with EBV and DYD as dependent variables did not reveal marked differences, the following results are presented based on DYD. As shown on Figure 1 for all the traits considered mBIC selects parsimonious models. In detail, based on mBIC, for milk and fat yields at both parities a model with the highest mBIC is the one containing additive and dominance effects of K232A (DGAT1). Consequently, the other polymorphisms have neither marginal nor epistatic effects. None of the polymorphisms considered (including DGAT1) affect protein yield. As expected from the lower penalty term, in comparison to mBIC, the original BIC points on models with higher parameterisation. Precisely:

a) for milk and fat yields at 1st parity as well as for protein yield at 2nd parity the model with $\mathbf{q}^T = \begin{bmatrix} a_{\mathrm{T945M}} & d_{\mathrm{T945M}} & a_{\mathrm{Y7F}} & d_{\mathrm{Y7F}} & a_{\mathrm{R25C}} & d_{\mathrm{R25C}} & a_{\mathrm{A80V}} & d_{\mathrm{A80V}} & a_{\mathrm{C/T}} & d_{\mathrm{C/T}} & a_{\mathrm{K232A}} & d_{\mathrm{K232A}} \\ a_{\mathrm{F16Y}} & d_{\mathrm{F16Y}} & a_{\mathrm{P35Q}} & d_{\mathrm{P35Q}} & a_{\mathrm{K468R}} & d_{\mathrm{K468R}} & \delta_{\mathrm{T945M,P35Q}} & & & & & \end{bmatrix}$ was selected,

b) for milk yield at 2nd parity, the same as in the case of mBIC, the model with $\mathbf{q}^T = \begin{bmatrix} a_{\mathrm{K232A}} & d_{\mathrm{K232A}} \end{bmatrix}$ was selected,

c) for protein yield at 1st parity the model with $\mathbf{q}^T = \begin{bmatrix} a_{\mathrm{T945M}} & d_{\mathrm{T945M}} & a_{\mathrm{Y7F}} & d_{\mathrm{Y7F}} & a_{\mathrm{R25C}} & d_{\mathrm{R25C}} & a_{\mathrm{A80V}} & d_{\mathrm{A80V}} & a_{\mathrm{C/T}} & d_{\mathrm{C/T}} & a_{\mathrm{K232A}} & d_{\mathrm{K232A}} \\ a_{\mathrm{F16Y}} & d_{\mathrm{F16Y}} & a_{\mathrm{P35Q}} & d_{\mathrm{P35Q}} & a_{\mathrm{K468R}} & d_{\mathrm{K468R}} & \delta_{\mathrm{K232A,P35Q}} & & & & & \end{bmatrix}$ was selected,

d) for fat yield at 2nd parity the model with $\mathbf{q}^T = \begin{bmatrix} a_{\text{T945M}} & d_{\text{T945M}} & a_{\text{Y7F}} & d_{\text{Y7F}} & a_{\text{R25C}} & d_{\text{R25C}} & a_{\text{A80V}} & d_{\text{A80V}} & a_{\text{C/T}} & d_{\text{C/T}} & a_{\text{K232A}} & d_{\text{K232A}} \\ a_{\text{F16Y}} & d_{\text{F16Y}} & a_{\text{P35Q}} & d_{\text{P35Q}} & a_{\text{K468R}} & d_{\text{K468R}} & \delta_{\text{K232A,Y7F}} \end{bmatrix}$ was selected.

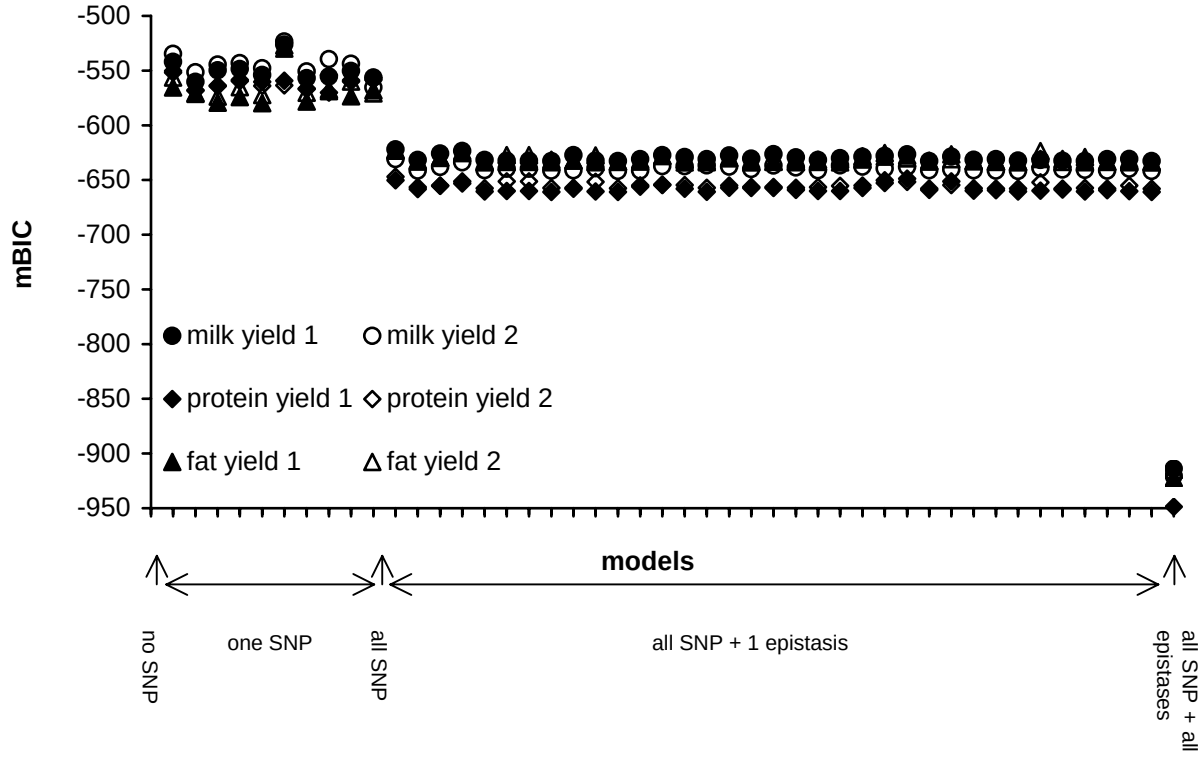

Figure 1 mBIC for different models and traits.

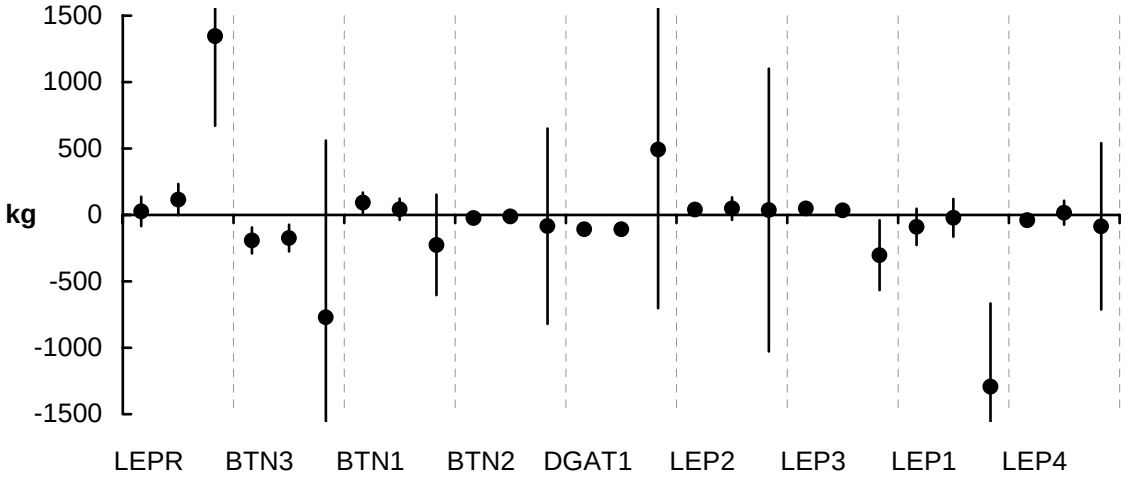

Figure 2 Estimates of additive effects of each SNP on milk yield at 1st lactation with corresponding 95% normal confidence intervals. For each SNP, from left to right, estimates from: a model with a single SNP, a model with all SNPs and a model with all SNPs and all possible pairwise additive-by-additive epistases are given.

Figure 2 presents estimates of additive effects of the polymorphisms on milk yield from 1st lactation under three different models: a model with a single SNP, a model with all SNPs and a model with all SNPs and all possible pairwise epistases. Estimates and their accuracy

(expressed by standard errors) do not differ much between both models without epistasis. When additive-by-additive epistasis is considered, the additional columns in the design matrix $\mathbf{X}_q$ result in highly elevated standard errors. The estimates of dominance effects show similar values and accuracy across models (Figure 3).

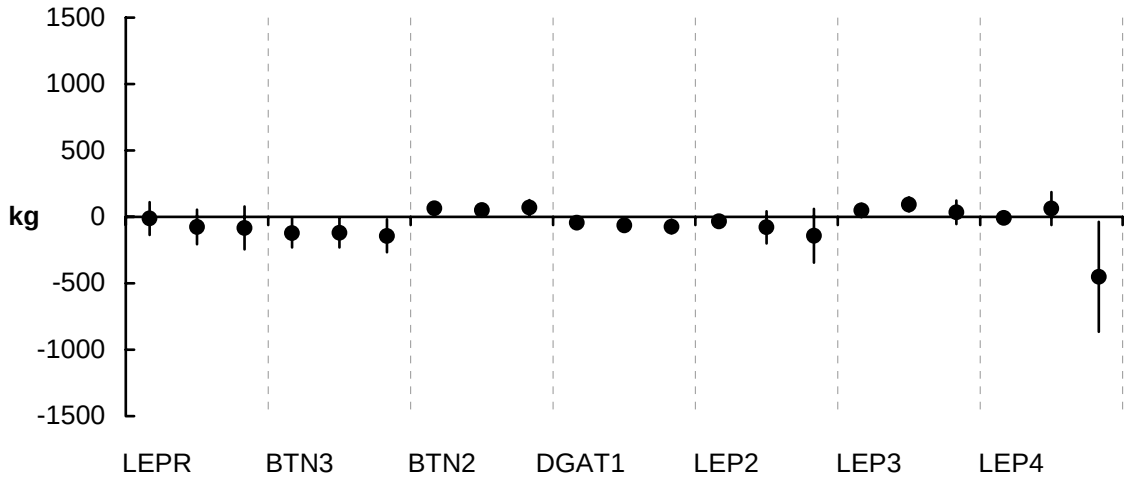

Figure 3 Estimates of dominance effects of each SNP on milk yield at 1st lactation with corresponding 95% normal confidence intervals. For each SNP, from left to right, estimates from: a model with a single SNP, a model with all SNPs and a model with all SNPs and all possible pairwise additive-by-additive epistases are given.

CONCLUSIONS

Presented results show no evidence of epistasis between the bytyrophilin gene, diacylglycerol acyltransferase 1 gene, leptin gene and leptin receptor. From all the polymorphisms considered the DGAT1 has a much larger effect on milk and fat yields then the other SNPs considered. In particular the additive effect of allele "GC/GC" amounts on average to 108.3 (109.0) kg milk and the dominance effect - 45.6 (63.6) kg milk, respectively for a model with DGAT1 as a single effect indicated as the best model by mBIC and for a model with main effects all nine SNPs.

Since analysed bulls have many daughters the difference between SNP effects estimated using EBV and DYD as a dependent variable are small. By definition, DYD represents bull's genetic merit and EBV represents genetic values of bulls plus ½ bull's sire and ½ bull's dam. While bull's EBV is estimated based on a large number of progeny the relative impact of the parental component is diminished, and thus DYD and EBV become similar. Still, in Table 2 it is shown that DYD have slightly higher variance than EBV. In contrast to EBV, DYD contain the residual component of the genetic evaluation model. Consequently, a variance of EBV is equivalent to the additive polygenic variance times reliability of the EBV, while a variance of DYD additionally contains a residual component. Considering the fact that candidate gene effects are likely to remain in the residual of the genetic evaluation model, DYD is a preferable dependent variable for gene detection.

Comparing both model selection criteria applied, i.e. mBIC and BIC it is evident that the letter selects much more highly parameterised models. However, as explained by Bogdan et al. (2004) the original version of BIC, as proposed by Schwarz, in case of multiple markers fails to keep the 5% type I error and tends to overestimate the number of QTL.

REFERENCES

Ashwell, M. S., Heyen, D. W., Sonstegard, T. S., Van Tassel, C. P., Da, Y., Van Raden, P. M., Ron, M., Weller, J. I. and H. A. Lewin. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J. Diary. Sci.* 87:468–475.

Ashwell, M.S., Rexroad Jr, C.E., Miller, R.H., Van Raden, P.M. and Y. Da. 1997. Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers. *Anim. Genet.* 28: 216–222.

Baierl, A., Bogdan, M. Frommlet, F. and A. Futschik. 2006. On locating multiple interacting quantitative trait loci in intercross designs. *Genetics* 173:1693-1703.

Bogdan, M., Gosh, J.K. and R. Doerge. 2004. Modifying the Schwarz Bayesian Information Criterion to locate multiple interacting quantitative trait loci. *Genetics* 167 :989-999.

Farese Jr, R.V., Cases, S. and S.J. Smith. 2000. Triglyceride synthesis: insights from the cloning of diacylglycerol acyltransferase. *Curr. Opin. Lipidol.* 11: 229-234.

Houseknecht, K.L., Baile, C.A., Matteri, R.L. and M.E. Spurlock. 1998. The biology of leptin: a review. *J. Anim Sci.* 76: 1405–1420.

Kao, C.H, and, Z.B. Zeng. 2002. Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics* 160:1243-1261.

Khatkar, M.S., Thomson, P.C., Tammen, I. and H.W. Raadsma. 2004. Quantitative trait locimapping in dairy cattle: Review and meta-analysis. *Genet. Se. Evol.* 36: 163-190.

Lindersson, M., Andersson-Eklund, L., de Koning, D.J., Lunden, A., Maki-Tanila, A. and L. Andersson. 1998. Mapping of serum amylase-1 and quantitative trait loci for milk production traits to cattle chromosome 4. *J. Dairy Sci.* 81: 1454-61.

Liu, Z., Reinhardt, F., Bunger, A., and R. Reents. 2004 Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. *J. Dairy Sci.* 87:1896-1907.

Ogg, S.L., Weldon, A.K., Dobbie, L., Smith, A.J.H. and I.H. Mather. 2004. Expression of butyrophilin (Btn1a1) in lactating mammary gland is essential for the regulated secretion of milk-lipid droplets. *Proc. Natl. Acad. Aci. USA* 101: 10084-10089.

Plante, Y., Gibson, J.P., Nadesalingam, J., Mehrabani-Yeganeh, H., Lefebvre, S., Vandervoort, G. and G.B Jansen. 2001. Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *J. Dairy Sci.* 84, 1515–1524.

Schwarz G. 1978 Estimating the dimension of a model. *Ann. Stat.* 6:461–464