Testing candidate gene effects on milk production traits in dairy cattle under various parameterizations and modes of inheritance

T. Suchocki,* J. Komisarek,† and J. Szyda*‡¹

*Department of Animal Genetics, Wrocław University of Environmental and Life Sciences, Kozuchowska 7, 51-631 Wrocław, Poland

†Department of Cattle Breeding and Milk Production, Poznań University of Life Sciences, Wojska Polskiego 71 A, 60-625 Poznań, Poland

‡Institute of Natural Sciences, Wrocław University of Life Sciences, Norwida 25, 50-375 Wrocław, Poland

ABSTRACT

The major objectives of this study were 1) to assess the statistical properties of models commonly used for the estimation of single nucleotide polymorphism (SNP) effects under the assumption of various modes of inheritance and various parameterizations of SNP genotypes using simulated data, and 2) to compare effects of the selected polymorphisms located within butyrophilin (*BTN1A1*), diacylglycerol acyltransferase 1 (*DGAT1*), leptin (*LEP*), and leptin receptor (*LEPR*) candidate genes on milk production traits using data from 2 dairy cattle breeds (190 Jersey cows and 475 Polish Holstein-Friesian cows). Simulation results showed that type I error and power were not dependent on the assumed parameterization, but differences were observed regarding confidence intervals of estimated SNP effects. In the presence of epistasis, correct confidence intervals for all (epistatic and nonepistatic) SNP and all modes of inheritance were provided only by the parameterization proposed by C. H. Kao and Z. B. Zeng in 2002. However, if no dominance effect was included in the model, confidence intervals for SNP effects were correct for all parameterizations. Results based on real data showed that for both breeds the additive effects of polymorphisms were generally similar, except for *LEPR*, which had a different allele associated with increased fat content in Holstein-Friesians than in Jerseys. In both breeds, *DGAT1* had the largest additive effect of the polymorphisms considered, but its effect on most milk traits was more pronounced in Jerseys than in Holstein-Friesians. Evidence of epistasis was found between *LEPR* and *DGAT1*, as well as between *LEPR* and *BTN1A1*, but only for milk content traits and only in the Holstein-Friesian breed. There was also more evidence for dominance in the Holstein-Friesian breed than in the Jersey breed.

Key words: candidate gene, epistasis, mixed model, statistical property

INTRODUCTION

For several years, SNP have been continuously gaining importance in revealing the genetic background of complex, quantitative characters. On one hand, in applications related to dairy cattle, a very large number of SNP can now be available for a single animal. This information can be used to predict the joint effect of (possibly) all additive gene effects responsible for a given trait, which is equivalent to the animal's breeding value (Goddard and Hayes, 2007). On the other hand, effects of single genes, the so-called candidate genes, marked by only several SNP, are estimated to identify causal mutations and decode traits' mode of inheritance. Both applications are of interest not only when determining the genetic background underlying quantitative traits but also for their use in marker assisted selection. Still, when estimating SNP effects, practical problems are encountered that are related mainly to the construction and choice of a statistical model. Among them there are differences in estimated SNP effects depending on the assumed mode of inheritance or on the effects parameterization. The most commonly considered modes of inheritance involve a purely additive mode, a mode also including dominance effects of SNP as well as most highly parameterized models incorporating epistatic relationships between SNP. Moreover, many different ways of parameterizing SNP genotypes are available (Cockerham, 1954; Kao and Zeng, 2002; Zeng et al., 2005; Álvarez-Castro and Carlborg, 2007). As a result, estimates of many candidate gene effects cannot be confirmed across different dairy breeds (Seefried, 2008) or even within a breed across populations (Ibeagha-Awemu et al., 2008).

In response to the above-mentioned problems, the main goals of our study were 1) to assess statistical properties of mixed linear models commonly used for the estimation of SNP effects under the assumption of various modes of inheritance and various parameteriza-

Received July 6, 2009.

Accepted February 10, 2010.

¹Corresponding author: joanna.szyda@up.wroc.pl

Figure 1. Graphical representation of simulated effects of SNP for various modes of inheritance when $a = 0.2\sigma_y^2$. Left diagram: additive mode of inheritance (Am); additive and dominance mode of inheritance (ADm). Right diagram: additive mode of inheritance with pairwise additive-by-additive epistasis.

tions of SNP genotypes using simulated data, and 2) to compare effects of selected candidate genes estimated by the same model using real data from 2 dairy cattle breeds.

MATERIALS AND METHODS

Simulated Data

The simulated data sets provided an approximation of the real data analyzed in the second phase of the study. Care was taken that the structure of both data sources defined by SNP were alike in linkage disequilibrium, pedigree relationship, sample size, and the distribution of dependent variables. In particular, the simulated data comprised 250 (500) cows belonging to 5 paternal half sib families, resulting from mating each of 5 unrelated sires to 50 (100) unrelated dams. Genotypes and phenotypes were known for both generations. Genotypic information consisted of 5 unlinked SNP, each with a minor allele frequency of 0.3, representing candidate genes. Phenotypes were determined by 1) a residual effect with a zero mean and a variance $\sigma_e^2 = 0.7\sigma_y^2$, where σ_y^2 represents a phenotypic variance, which was assumed to be equal to 1; 2) an additive polygenic component with a 0 mean and a variance $\sigma_\alpha^2 = 0.3\sigma_y^2$ for individuals from the first generation and by a half of additive polygenic values of both parents plus a Mendelian sampling effect for the cows from the second generation; and 3) effects of the 5 SNP, which were dependent on the assumed mode of inheritance: a simulated additive effect of a single SNP amounted to $0.2\sigma_y^2$ or $0.1\sigma_y^2$, a dominance effect to $0.1\sigma_y^2$, and an epistatic effect of $0.05\sigma_y^2$ defined only for an additive-by-additive relationship between 2 SNP (Figure 1). For each set of the simulated conditions, 1,000 data sets were generated. The data sets simulated under the null hypothesis

were used to assess the type I error rate, whereas the data sets simulated under the alternative hypothesis were used to calculate the power. Both quantities were defined in terms of finding a significant SNP effect. Associated hypotheses and test statistics are described below.

Real Data

The 2 real data sets corresponded to 190 Jersey dairy cows and 475 Polish Holstein-Friesian dairy cows originating from the same herd. The cows were daughters of 20 and 112 sires, respectively, with the sire half-sib family sizes varying between 1 and 55 and 1 and 53 for Jerseys and Holstein-Friesians, respectively. Common features of both data sets were phenotype, environment, and genotype. Quantitative phenotypes represented dairy production traits: milk, fat, and protein yields as well as fat and protein contents, which are recorded on a routine basis and expressed as yields of 305-d lactation. Only phenotypes of the first lactations of each cow were considered. Table 1 gives basic descriptive statistics of the traits. The environment comprised similar management and geographical conditions because both data sets originated out of dairy production studs from the central-western part of Poland. Finally, the same 4 functional SNP polymorphisms, located within 4 unlinked candidate genes [the leptin receptor (*LEPR*), leptin (*LEP*), acyl-CoA:diacylglycerol acyltransferase 1 (*DGAT1*), and butyrophilin (*BTN1A1*)] were genotyped for both data sets. The choice of the candidate genes was not coincidental. Each of the genes is located on a different chromosome to ensure independent transmission of genotypes and consequently their orthogonality, which is required for some of the applied parameterizations to be valid (Cockerham, 1954). Moreover, based on functional similarity between the genes, quantified as semantic similarity between their

Table 1. Descriptive statistics for phenotypes of real data sets

305-d trait record	h^2	Jersey		Holstein-Friesian	
		Mean	SD	Mean	SD
Milk yield, kg	0.33	4,186.00	648.55	6,616.00	1,321.15
Fat yield, kg	0.29	231.80	40.42	275.90	51.43
Fat content, %	0.29	5.52	0.52	4.21	0.52
Protein yield, kg	0.29	159.20	21.84	222.10	43.35
Protein content, %	0.29	3.84	0.33	3.37	0.19

gene ontology (**GO**) terms following Jiang and Conrath (1997), epistasis between different genotypes of the genes may be expected. The summary of molecular and functional information for the genes is presented in Table 2. The highest semantic similarity was observed between *LEPR* and *BTN1A1* (7.3%), *LEPR* and *DGAT1* (7.2%), and *LEPR* and *LEP* (7.0%).

Table 3 summarizes information on the 4 SNP used to mark each of the candidate genes. Neither missing phenotypes nor genotypes were present in both data sets.

Statistical Model

The same form of a mixed linear model was used for the analysis of the simulated and both real data sets:

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

where $\mathbf{y}$ is a vector of trait values assuming $\mathbf{y} \sim N(\mathbf{X}\mathbf{q}, \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R})$ (and T indicates matrix transposition); $\mathbf{q}$ is a vector of fixed effects representing a general mean, calving year (for real data only), and SNP effects; $\boldsymbol{\alpha} \sim N(0, \mathbf{G})$ is a vector of random polygenic effects of animals with records assuming $\mathbf{G} = \mathbf{A}\sigma_\alpha^2$, with $\mathbf{A}$ representing additive polygenic relationships among individuals, estimated from a pedigree, and σ_α^2 being a component of the total additive genetic variance attributed to polygenes; $\mathbf{e} \sim N(0, \mathbf{R})$ is a vector of random errors, with $\mathbf{R} = \mathbf{I}\sigma_e^2$, where $\mathbf{I}$ represents an identity matrix and σ_e^2 denotes the error variance; and $\mathbf{X}$ and $\mathbf{Z}$ are corresponding design matrices. Both variance components were assumed as known (i.e., were not estimated) with $\sigma_\alpha^2 = 0.3\sigma_y^2$ and $\sigma_e^2 = 0.7\sigma_y^2$, whereas the

estimation of other parameters was based on solving the mixed model equations (Henderson, 1984):

$$\begin{bmatrix} \hat{\mathbf{q}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}.$$

Differences between applied models comprised the assumed mode of inheritance defined by $\mathbf{q}$ and parameterizations of SNP effects defined by $\mathbf{X}$. The following 4 modes of inheritance were considered:

1. The simplest, purely additive mode (**Am**), with the corresponding $\mathbf{q}$, given as

$$\mathbf{q}^T = \left[\mu, \sum_{i=1}^{N_{SNP}} a_i \right],$$

where μ is a general mean, N_{SNP} represents the number of SNP fitted (5 in simulations and 4 in real data sets) and a_i denotes the additive effect of SNP i .

2. An additive mode with pairwise additive-by-additive epistases (**AEm**), given as

$$\mathbf{q}^T = \left[\mu, \sum_{i=1}^{N_{SNP}} a_i, \sum_{i=1}^{N_{SNP}} \sum_{j=i}^{N_{SNP}} a_{ij} \right],$$

where a_{ij} denotes the effect of additive-by-additive epistasis between SNP i and j .

Table 2. Molecular and functional information on the candidate genes

Gene ¹	Nucleotide sequence code	Location	GO term for molecular function ²	Product
<i>LEPR</i>	NC_007301	BTA03	Cytokine receptor activity	Receptor, 1,165 aa
<i>LEP</i>	NC_007302	BTA04	Growth factor and hormone activity	Hormone, 167 aa
<i>DGAT1</i>	NC_007312	BTA14	Diacylglycerol o-acyltransferase activity	Enzyme, 489 aa
<i>BTN1A1</i>	NC_007324	BTA23	—	Signal protein, 526 aa

¹*LEPR* = leptin receptor; *LEP* = leptin; *DGAT1* = diacylglycerol acyltransferase 1; *BTN1A1* = butyrophilin.

²GO = gene ontology.

Table 3. Estimated genotype frequencies of SNP marking *BTN1A1*, *DGAT1*, *LEP*, and *LEPR* genes

Gene ¹	SNP	Reference	Genotype frequency		
			Jersey	Holstein-Friesian	<i>P</i> -value ²
<i>BTN1A1</i>	K468R	Taylor et al. (1996)	AA = 0.05; AG = 0.29; GG = 0.66	AA = 0.81; AG = 0.17; GG = 0.01	8.82 × 10 ⁻¹⁷
<i>DGAT1</i>	K232A	Grisart et al. (2002); Winter et al. (2002)	AA/AA = 0.42; AA/GC = 0.45; GC/GC = 0.13	AA/AA = 0.02; AA/GC = 0.26; GC/GC = 0.72	1.09 × 10 ⁻⁴⁰
<i>LEP</i>	A80V	Konfortov et al. (1999)	CC = 0.63; CT = 0.32; TT = 0.05	CC = 0.15; CT = 0.49; TT = 0.36	4.99 × 10 ⁻¹
<i>LEPR</i>	T945M	Liefers et al. (2004)	TT = 0.45; CT = 0.47; CC = 0.08	TT = 0.10; CT = 0.40; CC = 0.50	1.54 × 10 ⁻⁷

¹*BTN1A1* = butyrophilin; *DGAT1* = diacylglycerol acyltransferase 1; *LEP* = leptin; *LEPR* = leptin receptor.

²*P*-values correspond to a chi-squared test for testing the null hypothesis of equal allele frequencies in the 2 breeds.

3. An additive and dominance mode (**ADm**), given as

$$\mathbf{q}^T = \left[\mu, \sum_{i=1}^{N_{SNP}} a_i, \sum_{i=1}^{N_{SNP}} d_i \right],$$

where d_i denotes the dominance effect of SNP i .

4. The most highly parameterized, an additive and dominance mode with pairwise additive-by-additive epistases (**ADEm**), given as

$$\mathbf{q}^T = \left[\mu, \sum_{i=1}^{N_{SNP}} a_i, \sum_{i=1}^{N_{SNP}} d_i, \sum_{i=1}^{N_{SNP}} \sum_{j=i}^{N_{SNP}} a_{ij} \right].$$

The 5 different parameterizations of SNP effects considered in our study were defined by different design matrices **X**. The original parameterization proposed by Cockerham (1954) (**Cp**) for the partitioning of a phenotypic variance is expressed in terms of SNP allele (p_0 and p_1) and genotype frequencies (p_{11} , p_{01} , and p_{00}), with the corresponding elements of **X** given as

$$\mathbf{X} = \begin{bmatrix} 1 & 2p_0 & \frac{1}{p_{11}} \\ 1 & p_0 - p_1 & -\frac{2}{p_{01}} \\ 1 & -2p_1 & \frac{1}{p_{00}} \end{bmatrix},$$

where rows 1, 2, and 3 correspond to SNP genotypes 11, 01, and 00, respectively, whereas columns 1, 2, and 3 correspond to the general mean and additive and dominance SNP effects, respectively.

An F_∞ parameterization by Kao and Zeng (2002) (**KZ_∞p**) is based on Cockerham’s (1954) parameterization and a case when an F2 generation has undergone selfing for an infinite number of generations, resulting in maintaining equal frequencies of both homozygous genotypes and removing heterozygotes from the population, with **X** defined as

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

A parameterization proposed by Álvarez-Castro and Carlborg (2007) (**AC₁p**) is defined in terms of genotype frequencies:

$$\mathbf{X} = \begin{bmatrix} 1 & -p_{01} - 2p_{00} & -p_{01} \\ 1 & 1 - p_{01} - 2p_{00} & 1 - p_{01} \\ 1 & 2 - p_{01} - 2p_{00} & -p_{01} \end{bmatrix}.$$

This parameterization is equivalent to F_∞ under the aforementioned assumption (i.e., $p_{11} = p_{00} = \frac{1}{2}, p_{01} = 0$).

A parameterization of Álvarez-Castro and Carlborg (2007) (**AC₂p**) is defined for a general population:

$$\mathbf{X} = \begin{bmatrix} 1 & -p_{01} - 2p_{00} & \frac{2p_{01}p_{00}}{p_{11} + p_{00} - (p_{11} - p_{00})^2} \\ 1 & 1 - p_{01} - 2p_{00} & \frac{4p_{11}p_{00}}{p_{11} + p_{00} - (p_{11} - p_{00})^2} \\ 1 & 2 - p_{01} - 2p_{00} & \frac{2p_{11}p_{01}}{p_{11} + p_{00} - (p_{11} - p_{00})^2} \end{bmatrix}.$$

This parameterization is equivalent to Zeng et al. (2005) under the assumption of the Hardy-Weinberg equilibrium between allele and genotype frequencies.

For all the above models, the parameterization for pairwise additive-by-additive epistatic effects was defined as a product of columns of $\mathbf{X}$ corresponding to the additive effect of SNP. Note that for simulated data allele (p_0, p_1) as well as genotype (p_{00}, p_{01}, p_{11}), frequencies used in $\mathbf{X}$ were true (i.e., simulated values), whereas for real data those frequencies were estimated from Jersey and Holstein-Friesian data sets, respectively.

Hypotheses Testing

When considering statistical properties of these models, the primary interest was the ability to detect a true mode of inheritance, followed by an unbiased and accurate estimation of SNP effects. The former was realized by testing $H_0: \mathbf{q} \in \Theta_0$, when Θ_0 represents a restricted parameter space; for example:

$$\mathbf{q}^T = \left[\mu, \sum_{i=1}^{N_{SNP}} a_i = 0, \sum_{i=1}^{N_{SNP}} d_i = 0, \sum_{i=1}^{N_{SNP}} \sum_{j=i}^{N_{SNP}} a_{ij} = 0 \right].$$

It was done by the likelihood ratio test statistic

$$\lambda = -2 \ln \frac{L(\mathbf{q}_0)}{L(\mathbf{q})} = [\mathbf{y} - \mathbf{X}\mathbf{q}_0]^T \mathbf{V}^{-1} [\mathbf{y} - \mathbf{X}\mathbf{q}_0] - [\mathbf{y} - \mathbf{X}\mathbf{q}]^T \mathbf{V}^{-1} [\mathbf{y} - \mathbf{X}\mathbf{q}] \sim \chi_k^2,$$

where $\mathbf{q}_0$ and $\mathbf{q}$ represent a restricted and an unrestricted vector of fixed effects, respectively, $\mathbf{V}$ is the covariance of $\mathbf{y}$ defined as $\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ in the above model, and k represents the difference in the number of parameters between an unrestricted and a restricted model, as defined by corresponding modes of inheritance (Am, AEm, ADm, or ADEm). Tables 4 and 5 define $\mathbf{q}_0$ and $\mathbf{q}$ corresponding to particular tests.

The latter is assessed by calculating 95% confidence intervals for the estimates of SNP effects. Moreover, for the real data sets, the significance of each SNP effect was determined by the Wald test:

$$W = \frac{\hat{\mathbf{I}}\mathbf{q} - \mathbf{0}}{\sqrt{(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}}} \sim N(0,1)$$

for $H_0: \mathbf{I}\mathbf{q} = 0$ versus $H_0: \mathbf{I}\mathbf{q} \neq 0$.

Comparison of Modes of Inheritance

For the real data set, different modes of inheritance were compared using the Bayesian information criteria (BIC; Schwarz, 1978): $\text{BIC} = n \ln(L) + k \ln(n)$, where n represents the number of records, which was the same for all the compared models, and k (the number of estimated parameters) and L (the likelihood) represent quantities, which varied across modes of inheritance.

RESULTS

Analysis of Simulated Data

The empirical type I error rates of λ comparing various sets of restricted and unrestricted models are given in Table 4. For each comparison, the cutoff corresponds to the 0.05 quartile of the χ^2 distribution with an appropriate number of degrees of freedom. In most cases tests were somewhat conservative, with type I errors ranging between 0.022 and 0.050. No systematic differences between parameterizations and unrestricted model assumptions could be identified. Consequently, there were also no systematic differences between parameterizations regarding power (Table 5). In particular, power was the highest if the unrestricted models were compared against

$$\mathbf{q}^T = \left[\mu, \sum_{i=1}^{N_{SNP}} a_i = 0, \sum_{i=1}^{N_{SNP}} d_i = 0, \sum_{i=1}^{N_{SNP}} \sum_{j=i}^{N_{SNP}} a_{ij} = 0 \right]$$

and was much higher for $a = 0.2$, ranging between 0.620 and 0.910, than for $a = 0.1$, ranging between 0.135 and 0.310. Including dominance in the unrestricted model caused the reduction of power. When a larger data set of 500 individuals was available, power (logically) increased. In particular, we observed an approximately linear increase in power with the increase of the sample size. For example, by increasing the sample size from 250 to 500 power, $\text{KZ}_{\infty p}$ increased from 0.279 to 0.558 (Am), 0.293 to 0.603 (AEm), 0.166 to 0.395 (ADm), and 0.152 to 0.415 (ADEm) when testing against the restricted model without SNP effects. The relatively higher increase was observed for Cp. Still, there were no marked differences in power between different parameterizations. However, for all data sets and effects sizes considered in the simulation there was not enough power to detect dominance or epistasis. The power for detecting a dominance model was practically on the assumed level of type I error.

Another aspect considered in the simulations was the coverage of the true effect by the confidence in-

Table 4. Type I error rate of the likelihood ratio test for the 0.05 quantile of the chi-squared distribution, calculated for 250 individuals and based on 1,000 simulated replications^{1,2}

True mode of inheritance	Unrestricted mode of inheritance				Parameterization ³
	Am	AEm	ADm	ADEm	
No genetic effect	0.038	0.026	0.039	0.033	Cp
	0.048	0.047	0.036	0.037	KZ _∞ P
	0.034	0.034	0.031	0.028	AC ₁ P
	0.028	0.035	0.024	0.027	AC ₂ P
Am	—	0.037	0.028	0.032	Cp
	—	0.043	0.032	0.036	KZ _∞ P
	—	0.043	0.039	0.036	
	—	0.034	0.033	0.027	AC ₁ P
	—	0.038	0.032	0.033	
	—	0.041	0.023	0.026	AC ₂ P
	—	0.043	0.037	0.039	
	—	0.041	0.044	0.035	
AEm	—	—	—	0.045	Cp
	—	—	—	0.026	KZ _∞ P
	—	—	—	0.032	
	—	—	—	0.026	AC ₁ P
	—	—	—	0.023	
	—	—	—	0.022	AC ₂ P
	—	—	—	0.037	
	—	—	—	0.032	
ADm	—	—	—	0.028	Cp
	—	—	—	0.034	KZ _∞ P
	—	—	—	0.040	
	—	—	—	0.029	AC ₁ P
	—	—	—	0.032	
	—	—	—	0.045	AC ₂ P
	—	—	—	0.030	
	—	—	—	0.050	

¹Within each parameterization, the first row corresponds to $a = 0.2$ and the second row to $a = 0.1$. The true (simulated) mode of inheritance corresponds to $\mathbf{q}_0$.

²Am = additive mode; AEm = additive mode with additive-by-additive epistases; ADm = additive and dominance mode; ADEm = additive and dominance mode with pairwise additive-by-additive epistases.

³See main text for parameterizations.

tervals estimated for different parameterizations and models, which is summarized in Figure 2. If a given SNP was not involved in an epistatic relationship, its additive affects were correctly estimated, regardless of the parameterization and model used. But in the case of epistatic SNP, correct 95% confidence intervals for all the models were provided only by KZ_∞p parameterizations, with the coverage ranging between 93.32 and 96.14%. Confidence intervals of epistatic SNP based on Cp, AC₁p, and AC₂p were correct only if there was no dominance in the model. Among the estimates with correct confidence intervals, neither increased effect (from 0.1 to 0.2) nor sample size (from 250 to 500) resulted in further improvement of the accuracy of estimation.

Candidate Gene Effects Estimation for Real Data

A comparison of estimates of the additive effect of *BTN1A1*, *DGAT1*, *LEP*, and *LEPR* for the 2 breeds,

estimated using AC₁p parameterization with Am and ADEm modes, is given in Figure 3. Detailed estimates of additive, dominance, and epistatic effects of the 4 candidate genes in Holstein-Friesians and Jerseys under various modes of inheritance and parameterizations are given in Table 6.

Summarizing the results of additive effects of particular genes, we observed that *LEPR* significantly affected fat and protein contents. Interestingly, for fat content the additive effects had opposite signs in Holstein-Friesians (allele C increases content) and in Jerseys (allele C decreases content), whereas for protein content allele C increased content in both breeds unless dominance was included for Jerseys. At *BTN1A1*, allele G decreased protein content in both breeds. However, this effect was significant only in Holstein-Friesians because of much smaller confidence intervals than those in Jerseys. The gene *DGAT1* (allele AA) significantly decreased milk and protein yields (for some models)

Table 5. Power of the likelihood ratio test for the 0.05 quantile of the chi-squared distribution, based on 1,000 simulated replications^{1,2}

Restricted mode of inheritance	Unrestricted mode of inheritance				Parameterization ³
	Am	AEm	ADm	ADEm	
No genetic effect	0.904	0.910	0.655	0.658	Cp
	0.279	0.288	0.156	0.135	
	0.558	0.621	0.351	0.430	
	0.907	0.899	0.649	0.682	KZ _∞ P
	0.279	0.293	0.166	0.152	
	0.558	0.603	0.395	0.415	
	0.901	0.888	0.641	0.649	AC ₁ P
	0.270	0.309	0.153	0.158	
	0.569	0.592	0.358	0.430	
	0.900	0.906	0.620	0.667	AC ₂ P
	0.264	0.292	0.163	0.167	
	0.547	0.562	0.380	0.402	
Am	—	0.036	0.125	0.115	Cp
	—	0.024	0.152	0.101	
	—	0.041	0.261	0.204	
	—	0.043	0.138	0.130	KZ _∞ P
	—	0.041	0.123	0.102	
	—	0.051	0.193	0.206	
	—	0.032	0.122	0.126	AC ₁ P
	—	0.037	0.143	0.113	
	—	0.047	0.113	0.126	
	—	0.042	0.138	0.105	AC ₂ P
	—	0.045	0.276	0.235	
	—	—	—	—	
AEm	—	—	—	0.125	Cp
	—	—	—	0.106	
	—	—	—	0.216	
	—	—	—	0.136	KZ _∞ P
	—	—	—	0.103	
	—	—	—	0.236	
	—	—	—	0.130	AC ₁ P
	—	—	—	0.118	
	—	—	—	0.246	
	—	—	—	0.121	AC ₂ P
	—	—	—	0.108	
	—	—	—	0.230	
ADm	—	—	—	0.046	Cp
	—	—	—	0.039	
	—	—	—	0.038	
	—	—	—	0.041	KZ _∞ P
	—	—	—	0.039	
	—	—	—	0.049	
	—	—	—	0.043	AC ₁ P
	—	—	—	0.035	
	—	—	—	0.050	
	—	—	—	0.053	AC ₂ P
	—	—	—	0.030	
	—	—	—	0.043	

¹In each parameterization, the first row corresponds to $a = 0.2$ and 250 individuals, the second row to $a = 0.1$ and 250 individuals, and the third row to $a = 0.1$ and 500 individuals. The true (simulated) mode of inheritance is ADEm.

²Am = additive mode; AEm = additive mode with additive-by-additive epistases; ADm = additive and dominance mode; ADEm = additive and dominance mode with pairwise additive-by-additive epistases.

³See main text for parameterizations.

and increased fat yield and fat and protein contents. However, the increasing effect of AA on protein and fat concentrations was more pronounced in Jerseys than in Holstein-Friesians, even though the increase in fat yield was higher in the latter breed. Consequently, the higher

concentration effect estimated for Jerseys was attributable to the fact that AA strongly influenced decrease of milk yield in this breed. For *LEP*, allele C (depending on the model) had a significant effect, resulting in an increase of milk yield and fat yield in Jerseys as well as

Figure 2. Coverage of the simulated additive SNP effect by the estimated 95% confidence intervals for the 5 simulated SNP. Black circles represent SNP with no epistatic relationship; gray circles represent SNP with epistatic relationship. The samples of 250 individuals were generated under the true (simulated) ADEm mode. (A) Graphs relate to the additive SNP effect of $0.1\sigma_y^2$; (B) graphs relate to the additive SNP effect of $0.2\sigma_y^2$. Am = additive mode; AEm = additive-by-additive epistases; ADm = additive and dominance mode; ADEm = additive and dominance mode with pairwise additive-by-additive epistases. See main text for definitions of parameterizations (Cp, $KZ_{\infty}p$, AC_1p , and AC_2p).

Figure 3. Estimates and 95% confidence intervals of additive effects of SNP marking butyrophilin (*BTN1A1*), *DGAT1* (diacylglycerol acyltransferase 1), *LEP* (leptin), and *LEPR* (leptin receptor) genes. For each gene, the 2 dots in the same color represent effects based on additive mode of inheritance (left dot) as well as on additive and dominance mode with pairwise additive-by-additive epistases (right dot) estimated with AC₁p parameterization (see main text for definition). Grey dots represent Holstein-Friesian cows; black dots represent Jersey cows. Estimates were multiplied by (−1) to retain correspondence with other parameterizations.

a nonsignificant effect on the increase in protein yield. At the same time, this allele reduced protein content in Jerseys, meaning that an increase in milk yield was more profound than a simultaneous increase in protein yield. In Holstein-Friesians the effect of *LEP* was nonsignificant, with point estimates of additive effects being close to 0 for all traits.

Estimates of the dominance effect and its significance differed considerably across breeds, models, and parameterizations. In general, the Holstein-Friesian data set provided more evidence of dominance than the Jersey data set. In particular, a significant dominance effect of *DGAT1* on milk yield was estimated for both Holstein-Friesians ($P < 0.00347$ for ADm; $P < 0.00829$

Table 6. Estimates of additive effects of SNP marking *BTN1A1*, *DGAT1*, *LEP*, and *LEPR* genes^{1,2}

Effect	Holstein-Friesian				Jersey			
	Cp	KZ _∞ P	AC ₁ P	AC ₂ P	Cp	KZ _∞ P	AC ₁ P	AC ₂ P
Milk yield, kg								
Am								
<i>a</i> _{LEPR}	-236.54	-235.92	236.54	—	40.21	40.18	-39.78	—
<i>a</i> _{BTN1A1}	-2.32	-2.69	2.32	—	102.77	101.43	-101.72	—
<i>a</i> _{DGAT1}	-263.48	-263.49	263.48	—	-342.78	-339.67	339.33	—
<i>a</i> _{LEP}	-76.66	-76.81	76.66	—	142.54	140.80	-141.09	—
ADm								
<i>a</i> _{LEPR}	-225.21	-131.90	132.05	219.77	52.37	25.95	-24.86	-33.48
<i>a</i> _{BTN1A1}	-3.45	15.30	-15.37	11.02	105.38	74.62	-74.99	-97.16
<i>a</i> _{DGAT1}	-266.70	-277.02	277.13	253.58	-357.08	-303.40	302.47	290.82
<i>a</i> _{LEP}	-72.76	-23.75	23.47	67.24	139.60	214.35	-214.89	-135.75
Aem ³								
<i>a</i> _{LEPR}	-244.01	-158.22	244.01	—	50.90	17.96	-50.95	—
<i>a</i> _{BTN1A1}	-3.17	-37.75	3.17	—	115.39	138.79	-113.66	—
<i>a</i> _{DGAT1}	-269.93	-141.16	269.93	—	-350.30	-390.29	346.82	—
<i>a</i> _{LEP}	-60.71	-58.34	60.71	—	123.58	189.64	-121.65	—
ADEm								
<i>a</i> _{LEPR}	-235.26	-60.68	137.27	224.48	63.52	72.68	-49.64	-41.84
<i>a</i> _{BTN1A1}	-5.63	25.99	-31.10	14.60	122.94	74.22	-95.05	-107.29
<i>a</i> _{DGAT1}	-286.60	-147.71	291.43	260.81	-359.65	-344.39	294.32	299.30
<i>a</i> _{LEP}	-45.66	-15.65	2.14	52.73	135.34	2,798.34	-192.28	-117.08
Fat yield, kg								
Am								
<i>a</i> _{LEPR}	-1.63	-1.61	1.63	—	3.37	3.37	-3.34	—
<i>a</i> _{BTN1A1}	-3.78	-3.80	3.78	—	7.30	7.20	-7.23	—
<i>a</i> _{DGAT1}	9.05	9.05	-9.05	—	1.97	1.91	-1.96	—
<i>a</i> _{LEP}	-1.73	-1.73	1.73	—	7.96	7.86	-7.88	—
ADm								
<i>a</i> _{LEPR}	-0.86	-0.18	0.18	1.86	3.80	-1.45	1.55	-2.66
<i>a</i> _{BTN1A1}	-4.02	-1.42	1.41	3.85	7.52	6.46	-6.50	-6.80
<i>a</i> _{DGAT1}	9.04	7.44	-7.43	-8.63	1.07	2.79	-2.88	-1.73
<i>a</i> _{LEP}	-1.35	-2.58	2.57	1.56	7.84	12.03	-12.07	-7.45
AEm ³								
<i>a</i> _{LEPR}	-2.25	-2.58	2.25	—	3.27	4.45	-3.27	—
<i>a</i> _{BTN1A1}	-4.03	-3.24	4.03	—	6.98	-0.03	-6.89	—
<i>a</i> _{DGAT1}	7.18	2.32	-7.18	—	1.06	-2.08	-1.05	—
<i>a</i> _{LEP}	-0.50	-0.64	0.50	—	7.24	9.86	-7.14	—
ADEm								
<i>a</i> _{LEPR}	-1.34	0.14	0.51	2.40	4.01	-0.05	-0.17	-2.59
<i>a</i> _{BTN1A1}	-4.29	-0.46	0.93	4.24	7.59	-1.10	-5.45	-6.49
<i>a</i> _{DGAT1}	6.58	1.35	-5.42	-6.83	0.30	-2.53	-1.89	-0.94
<i>a</i> _{LEP}	0.35	0.18	0.76	0.42	7.91	14.60	-10.57	-6.82
Fat content, %								
Am								
<i>a</i> _{LEPR}	0.12	0.12	-0.12	—	-0.10	-0.10	0.10	—
<i>a</i> _{BTN1A1}	-0.04	-0.04	0.04	—	0.07	0.07	-0.07	—
<i>a</i> _{DGAT1}	0.28	0.28	-0.28	—	0.45	0.44	-0.44	—
<i>a</i> _{LEP}	0.02	0.02	-0.02	—	-0.07	-0.07	0.06	—
ADm								
<i>a</i> _{LEPR}	0.12	0.08	-0.08	-0.11	-0.10	-0.21	0.21	0.09
<i>a</i> _{BTN1A1}	-0.04	-0.03	0.03	0.04	0.08	0.07	-0.07	-0.07
<i>a</i> _{DGAT1}	0.28	0.27	-0.27	-0.27	0.44	0.41	-0.41	-0.38
<i>a</i> _{LEP}	0.03	-0.02	0.02	-0.02	-0.06	-0.04	0.04	0.06
AEm ³								
<i>a</i> _{LEPR}	0.12	0.07	-0.12	—	-0.11	-0.02	0.11	—
<i>a</i> _{BTN1A1}	-0.04	-0.01	0.04	—	0.03	-0.06	-0.03	—
<i>a</i> _{DGAT1}	0.26	0.11	-0.26	—	0.44	0.43	-0.43	—
<i>a</i> _{LEP}	0.03	0.04	-0.03	—	-0.06	0.19	0.06	—
ADEm								
<i>a</i> _{LEPR}	0.12	0.05	-0.08	-0.10	-0.10	-0.10	0.18	0.10
<i>a</i> _{BTN1A1}	-0.05	-0.01	0.03	0.04	0.04	0.04	-0.01	-0.03
<i>a</i> _{DGAT1}	0.26	0.10	—	-0.25	0.43	0.43	-0.39	-0.37
<i>a</i> _{LEP}	0.03	0.02	0.01	-0.03	-0.06	-0.06	0.05	0.06

Continued

Table 6 (Continued). Estimates of additive effects of SNP marking *BTN1A1*, *DGAT1*, *LEP*, and *LEPR* genes^{1,2}

Effect	Holstein-Friesian				Jersey			
	Cp	KZ _∞ p	AC ₁ p	AC ₂ p	Cp	KZ _∞ p	AC ₁ p	AC ₂ p
Protein yield, kg								
Am								
<i>a_{LEPR}</i>	−4.88	−4.85	4.88	—	0.35	0.36	−0.35	—
<i>a_{BTN1A1}</i>	−2.73	−2.75	2.73	—	2.82	2.78	−2.79	—
<i>a_{DGAT1}</i>	−4.56	−4.56	4.56	—	−7.67	−7.60	7.58	—
<i>a_{LEP}</i>	−1.66	−1.67	1.66	—	2.90	2.86	−2.87	—
ADm								
<i>a_{LEPR}</i>	−4.49	−2.47	2.48	4.68	0.72	0.67	−0.62	−0.35
<i>a_{BTN1A1}</i>	−2.85	−0.97	0.97	2.78	2.79	2.19	−2.21	−2.68
<i>a_{DGAT1}</i>	−4.60	−5.00	5.00	4.40	−8.06	−5.18	5.14	6.37
<i>a_{LEP}</i>	−1.51	−0.67	0.66	1.45	2.73	5.34	−5.36	−2.87
Aem ³								
<i>a_{LEPR}</i>	−5.14	−4.19	5.14	—	0.70	0.71	−0.70	—
<i>a_{BTN1A1}</i>	−2.67	−2.56	2.67	—	3.25	2.06	−3.20	—
<i>a_{DGAT1}</i>	−4.50	−2.48	4.50	—	−7.55	−10.02	7.48	—
<i>a_{LEP}</i>	−1.38	−1.73	1.38	—	2.93	7.23	−2.87	—
ADEm								
<i>a_{LEPR}</i>	−4.74	−1.17	2.58	4.89	1.03	0.65	−1.10	−0.62
<i>a_{BTN1A1}</i>	−2.82	−0.22	0.69	2.76	3.42	1.00	−2.68	−3.03
<i>a_{DGAT1}</i>	−4.91	−2.61	5.02	4.36	−7.90	−8.40	5.31	6.35
<i>a_{LEP}</i>	−0.97	−0.52	0.35	1.20	3.34	9.45	−5.34	−2.89
Protein content, %								
Am								
<i>a_{LEPR}</i>	0.06	0.06	−0.06	—	0.03	0.03	−0.03	—
<i>a_{BTN1A1}</i>	−0.04	−0.04	0.04	—	−0.03	−0.03	0.03	—
<i>a_{DGAT1}</i>	0.05	0.05	−0.05	—	0.16	0.16	−0.16	—
<i>a_{LEP}</i>	0.01	0.01	−0.01	—	−0.06	−0.06	0.06	—
ADm								
<i>a_{LEPR}</i>	0.06	0.04	−0.03	−0.05	0.03	−0.01	0.01	−0.02
<i>a_{BTN1A1}</i>	−0.04	−0.02	0.02	0.04	−0.03	0.00	0.00	0.02
<i>a_{DGAT1}</i>	0.05	0.05	−0.05	−0.04	0.16	0.18	−0.18	−0.14
<i>a_{LEP}</i>	0.01	0.00	0.00	−0.01	−0.06	−0.06	0.06	0.05
Aem ³								
<i>a_{LEPR}</i>	0.06	0.03	−0.06	—	0.02	0.06	−0.02	—
<i>a_{BTN1A1}</i>	−0.04	−0.01	0.04	—	−0.03	−0.09	0.03	—
<i>a_{DGAT1}</i>	0.06	0.03	−0.06	—	0.17	0.17	0.17	—
<i>a_{LEP}</i>	0.01	0.01	−0.01	—	−0.04	−0.01	0.04	—
ADEm								
<i>a_{LEPR}</i>	0.06	0.02	−0.03	−0.05	0.02	−0.03	0.01	−0.02
<i>a_{BTN1A1}</i>	−0.04	−0.01	0.03	0.04	−0.03	−0.05	0.02	0.03
<i>a_{DGAT1}</i>	0.06	0.03	−0.06	−0.05	0.16	0.15	−0.16	−0.14
<i>a_{LEP}</i>	0.01	0.00	0.01	−0.01	−0.04	−0.04	0.04	0.04

¹Effects significant at the 5% level are marked in bold.²Am = additive mode; AEm = additive mode with additive-by-additive epistases; ADm = additive and dominance mode; ADEm = additive and dominance mode with pairwise additive-by-additive epistases. *BTN1A1* = butyrophilin; *DGAT1* = diacylglycerol acyltransferase 1; *LEP* = leptin; *LEPR* = leptin receptor. See main text for parameterizations (Cp, KZ_∞p, AC₁p, AC₂p).³This is the most probable mode of inheritance.

for ADEm) and Jerseys ($P < 0.00025$ for ADm; $P < 0.01247$ for ADEm), but based only on AC₂p. Similarly, under ADm ($P < 0.01230$) and ADEm ($P < 0.05514$), AC₂p gave some evidence of the dominance effect of *DGAT1* for fat yield in Holstein-Friesians, whereas the effect in Jerseys was on the border of significance. In contrast, the dominance effect of *DGAT1* on protein yield was significant only in Jerseys ($P < 0.01757$ using ADm and AC₂p). Furthermore, using ADm and ADEm with AC₂p, a highly significant dominance effect of

DGAT1 was observed for fat and protein contents for both breeds. For *LEPR*, weak evidence of dominance was detected by Cp in both breeds for fat content, but it disappeared after an epistatic effect was included into the model. In Holstein-Friesians, *LEPR* also had a significant dominance effect on protein yield, reported by ADm and ADEm and most of the parameterizations. Also, the dominance effect of *LEP* on both content traits was significant only in Holstein-Friesians, but the level of significance was not very high, ranging

between 0.01094 for fat content (ADm) and 0.056703 for protein content (ADEm). The dominance effect of *BTN1A1* showed only a borderline significance for protein content in Holstein-Friesians, albeit reported only by ADm with all parameterizations except for Cp.

Significant evidence of epistasis was observed only in Holstein-Friesians and only for both content traits. For fat content, a significant additive-by-additive epistasis between *LEPR* and *DGAT1* was estimated by both AEm with Cp and AC_{1p} ($P < 0.04534$) and by ADEm, but only when using Cp ($P < 0.01398$). For protein content, there was strong evidence of additive-by-additive epistasis between *LEPR* and *BTN1A1*, estimated by all parameterizations for AEm (P -values ranging between 0.000773 and 0.008585) and for Cp and AC_{2p} for ADEm ($P < 0.000643$ and $P < 0.002936$, respectively). For both traits, the incorporation of a dominance effect reduced the evidence for additive-by-additive epistasis. Still, for both breeds and all traits, AEm was selected by BIC as the most probable mode of inheritance.

DISCUSSION

Similarities between parameterizations regarding type I error and power could have been expected because the main differences between them comprise the definition of SNP effects. Their statistical properties can be influenced only by the fact that Cp, AC_{1p}, and AC_{2p} use estimates of allele frequencies, which are attributed to sampling error expressed as $\text{var}[\hat{p}_0] = \frac{p_0 p_1}{N}$, whereas KZ_{∞p} uses assumed (not estimated) allele frequencies, which do not depend on the sample size but may considerably differ from the true frequencies in the population analyzed. Yet, our simulation results showed that type I error and power of the tests did not depend on the parameterization.

Notable differences between parameterizations were observed regarding SNP effect estimation, showing that the only parameterization providing unbiased additive effects of epistatic SNP was KZ_{∞p}, whereas the other parameterizations considered provided biased estimates if dominance was included in a model. A similar analysis was also presented by Zeng et al. (2005), showing that unbiased estimates of additive SNP effects were guaranteed only if all effects of the true mode of inheritance were present in the statistical model. Although it is not possible to directly compare both studies because they assume different structures of the simulated data (residual variance and a relationship structure), our results were different by showing that KZ_{∞p} always provided unbiased estimates, even if the statistical model was misspecified. Both analyses are concordant in showing

that parameterizations involving SNP allele frequency, under certain conditions, result in biased estimates.

Following Zeng et al. (2005) and assuming ADm for one locus case, the estimator of an additive SNP effect is given as $\hat{a} = D_{21}G_{00} + D_{22}G_{01} + D_{23}G_{11}$, where D_{ij} represents the ij th element of the inverse of the $\mathbf{X}$ matrix for a given parameterization and G_{ij} is the genotypic mean for genotype ij . For KZ_{∞p}, we obtain the correct estimator of

$$\hat{a} = \frac{1}{2}G_{00} + 0G_{01} - \frac{1}{2}G_{11} = \frac{1}{2}(G_{00} - G_{11}).$$

For AC_{1p}, the estimator is given as

$$\hat{a} = -\frac{1}{2}G_{00} + 0G_{01} + \frac{1}{2}G_{11} = \frac{1}{2}(G_{11} - G_{00}),$$

which is also correct, only with the opposite sign. The corresponding estimator under Cp has a more complicated form of

$$\hat{a} = -\frac{p_{00}p_{01}p_{11}}{p_{01}p_{11} + 4p_{00}p_{11} + p_{00}p_{01}} \times \left[\left(-\frac{1}{p_{00}} - \frac{2}{p_{01}} \right) G_{00} + \left(\frac{1}{p_{00}} - \frac{1}{p_{11}} \right) G_{01} + \left(\frac{2}{p_{01}} + \frac{1}{p_{11}} \right) G_{11} \right],$$

which is a function of genotype frequencies. Under the assumption of $p_{00} = p_{11} = \frac{1}{4}$ and $p_{01} = \frac{1}{2}$ it is equal to $\hat{a} = \frac{1}{2}(G_{00} - G_{11})$, which is correct. Similarly, for AC_{2p}, it is a function of genotype frequencies and is given as

$$\frac{-p_{01}p_{11}(G_{00} - G_{01}) + 2p_{00}p_{11}(G_{00} - G_{11}) - p_{00}p_{01}(G_{01} - G_{11})}{p_{01}p_{11} - 4p_{00}p_{11} + p_{00}p_{01}}.$$

However, it is undefined for $p_{00} = p_{11} = \frac{1}{4}$ and $p_{01} = \frac{1}{2}$ because the denominator is equal to 0. Moreover, when a 2 loci model with epistasis is concerned, which is parameterized by 2 loci genotypic values ($G_{i'j'}$, where the subscript denotes genotype i' at locus i and genotype j' at locus j), the estimator of an additive effect of locus j provided by AC_{1p} is dependent on the allele frequencies at locus i :

$$\hat{a}_j = \frac{1}{2} \left[\frac{p_{\bullet-11}(G_{00-11} - G_{00-00}) + p_{\bullet-01}(G_{01-11} - G_{01-00})}{p_{\bullet-00}(G_{11-11} - G_{11-00})} \right].$$

Because in our parameterization true (i.e., assumed in simulation) genotype frequencies were used, which could differ from the frequencies observed in the simulated data sets, this might cause estimation bias as shown in Figure 2.

The 2 real data sets analyzed in the paper provided a unique opportunity to compare effects of the same genes in 2 breeds, which were subjected to the same management and selection goals. Still, some differences could be observed between both breeds. In particular, evidence of epistasis was found only in Holstein-Friesian cows. This either may have a biological basis or may result from the fact that the data set for Jerseys was much smaller and thus provided less power to detect epistasis. It is noteworthy that for loci with one of the homozygous genotype classes equal to 0, obtained estimates of epistasis did not represent true epistasis in the genetic sense. This is given as (Zeng et al., 2005)

$$\frac{(p_{11-11}G_{00-11} - 2p_{00-11}G_{11-00}) - (p_{00-11}G_{00-11} - p_{00-00}G_{00-00})}{p_{11-11} + p_{11-00} + p_{00-11} + p_{00-00}}.$$

There was also more evidence for dominance in Holstein-Friesians than in Jerseys; the possible reasons are indicated above. Significant differences in candidate gene effect estimates between dairy cattle breeds (including German Holstein-Friesian) have also been observed by Seefried (2008).

Among the 4 polymorphisms considered in this study, *DGAT1* K232A has been the most intensively investigated in cattle. In all populations examined so far, the estimated effects on milk traits were generally consistent with respect to their direction, with the lysine encoding AA variant increasing fat and protein contents as well as fat yield but reducing milk and protein yields (Grisart et al., 2002; Thaller et al., 2003; Bennewitz et al., 2004). The magnitude of K232A effects, however, differed across studies, both within and between dairy cattle breeds. Values obtained for Holstein-Friesian cows in this study are very similar to those previously published by Bennewitz et al. (2004) and Kaupe et al. (2007), but are higher than those reported by Grisart et al. (2002) for the Dutch population and by Thaller et al. (2003). Compared with results for the New Zealand and Swedish Holstein-Friesians, the estimated effects were smaller for fat and protein concentrations and generally larger for the other traits (Grisart et al., 2002; Näslund et al., 2008).

Several authors suggested differences in size of *DGAT1* effects between Holstein-Friesian cattle and other dairy breeds, such as Jersey, Ayrshire, Fleckvieh, Normande, Montbeliarde, and Angeln (Spelman et al., 2002; Thaller et al., 2003; Sanders et al., 2006). In this

study, the AA allele increased fat yield less markedly in Jersey cows than in Holstein-Friesian cows, which was consistent with results of Spelman et al. (2002). A similar tendency was also observed for Angeln and Normande cattle (Sanders et al., 2006; Gautier et al., 2007). However, effects estimated for the other traits were rather inconsistent across studies.

Differences in the magnitude of effects may be attributable to interactions with background genes, which may vary between populations. The gene *DGAT1* encodes the enzyme catalyzing the last step in triacylglycerol synthesis, and the enzyme variant including lysine was shown to have a greater activity compared with the variant with alanine (Grisart et al., 2004). According to Sanders et al. (2006), the decrease in the positive effect of the lysine (AA) allele on fat production in Jersey and Angeln breeds, characterized by a higher fat content in milk than that of Holstein-Friesians, may result from the limited availability of substrates for triacylglycerol synthesis. Background genes interacting with *DGAT1* might, therefore, be involved in diacylglycerol and fatty acyl-CoA metabolism or might compete for substrates with *DGAT1*. However, despite the lower fat percentage in milk, its yield in Holstein-Friesian cows is generally higher than in other breeds, which contradicts the above hypothesis. The insufficient supply of substrates for lipid synthesis in animals homozygous for the AA variant may, nevertheless, explain the dominance effect on milk fat content revealed in German Holstein-Friesians (Kuehn et al., 2007) and both in Polish Holstein-Friesian and Jersey cows (our study). On the other hand, some studies performed on *DGAT1* K232A polymorphism in cattle indicate its only additive mode of inheritance. The absence of dominance in New Zealand and Swedish Holstein-Friesians (529 and 96 cows, respectively), as well as in German Angelns (805 cows) and Swedish Reds (146 cows) was reported by Grisart et al. (2002), Sanders et al. (2006), and Näslund et al. (2008), respectively.

The epistasis identified for fat content between *LEPR* and *DGAT1* suggests a possible relationship between products of both genes. Although the underlying physiological mechanisms remain unclear, there is evidence that leptin, acting through its receptor, modulates some *DGAT1* functions (Chen et al., 2002b). Mice lacking *DGAT1* are characterized by an increased sensitivity of leptin (Chen et al. 2002a), which in turn may be involved in several pathways of milk lipid synthesis. In the presence of prolactin, leptin was shown to enhance fatty acid synthesis in the mammary gland (Feuermann et al., 2004). Leptin also suppresses the expression of *DGAT2*, the second mammalian CoA:diacylglycerol acyltransferase, playing, however, a less important role in milk production (Suzuki et al., 2005). The estimated

nonzero semantic similarities between some of the GO terms associated with products of *LEPR* and *DGAT1* also provided some indication of the functional basis underlying the epistatic effect.

The other polymorphisms analyzed in this paper, concerning *BTN1A1*, *LEP*, and *LEPR* genes, were not intensively investigated in cattle. Although some associations between *LEP* and milk production traits were found (Madeja et al., 2004; Komisarek et al., 2005), the obtained results differed between studies and the nonadditive effects were not estimated. Yet, considering the significant epistatic effect estimated on protein content between *LEPR* and *BTN1A1*, its functional representation is given by a nonzero semantic similarity (7.3%, the highest similarity among all considered GO terms) between receptor activity GO term associated with *BTN1A1* and protein binding GO term associated with *LEPR*. Both terms comprise numerous gene products (24,212 and 5,259, respectively).

Interesting results emerged when estimates of additive effects of the polymorphisms were compared between both data sets considered in our study. Although in general there were no differences in significance of the effects between Holstein-Friesians and Jerseys, some differences regarding the genetic background of protein synthesis are to be emphasized. In Jersey, *DGAT1* AA variant, in addition to decreasing milk yield, had a decreasing effect on protein yield, which was not observed in Holstein-Friesians. On the other hand, T945M and K468R, the 2 other polymorphisms located within *LEPR* and *BTN1A1* genes, respectively (Table 3), affected protein in the Holstein-Friesian breed but not in the Jersey breed.

ACKNOWLEDGMENTS

The study was financed by the Polish Ministry of Science and Higher Education, grant no. N N311 450434 and by Poznań University of Life Sciences grant no. 388/Z/32/W. We also acknowledge the anonymous reviewers, whose comments led to the improvement of the paper.

REFERENCES

- Álvarez-Castro, J. M., and Ö. Carlborg. 2007. A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics* 176:1151–1167.
- Bennewitz, J., N. Reinsch, S. Paul, C. Looft, B. Kaupe, C. Weimann, G. Erhardt, G. Thaller, Ch. Kühn, M. Schwerin, H. Thomsen, F. Reinhardt, R. Reents, and E. Kalm. 2004. The *DGAT1* K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. *J. Dairy Sci.* 87:431–442.
- Chen, H. C., J. S. Smith, Z. Ladha, D. R. Jensen, L. D. Ferreira, L. K. Pulawa, J. G. McGuire, R. E. Pitas, R. H. Eckel, and R. V. Farese Jr. 2002a. Increased insulin and leptin sensitivity in mice lacking acyl CoA:diacylglycerol acyltransferase 1. *J. Clin. Invest.* 109:1049–1055.
- Chen, H. C., J. S. Smith, B. Tow, P. M. Elias, and R. V. Farese Jr. 2002b. Leptin modulates the effects of acyl CoA:diacylglycerol acyltransferase deficiency on murine fur and sebaceous glands. *J. Clin. Invest.* 109:175–181.
- Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39:859–882.
- Feuermann, Y., S. J. Mabjeesh, and A. Shamay. 2004. Leptin affects prolactin action on milk protein and fat synthesis in the bovine mammary gland. *J. Dairy Sci.* 87:2941–2946.
- Gautier, M., A. Capitan, A. Fritz, A. Eggen, D. Boichard, and T. Druet. 2007. Characterization of the *DGAT1* K232A and variable number of tandem repeat polymorphisms in French dairy cattle. *J. Dairy Sci.* 90:2980–2988.
- Goddard, M. E., and B. J. Hayes. 2007. Genomic selection. *J. Anim. Breed. Genet.* 124:323–330.
- Grisart, B., Q. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. 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.* 12:222–231.
- Grisart, B., F. Farnir, L. Karim, N. Cambisano, J.-J. Kim, A. Kvasz, M. Mni, P. Simon, J. M. Frère, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the *DGAT1* K232A quantitative trait nucleotide in affecting milk yield and composition. *Proc. Natl. Acad. Sci. USA* 101:2398–2403.
- Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Ontario, Canada.
- Ibeagha-Awemu, E. M., P. Kgwatalala, and X. Zhao. 2008. A critical analysis of production-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mamm. Genome* 19:226–245.
- Jiang, J. J., and D. W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. Proceedings of International Conference Research on Computational Linguistics (ROCLING X), Taiwan.
- Kao, C. H., and Z. B. Zeng. 2002. Modelling epistasis of quantitative trait loci using Cockerham's model. *Genetics* 160:1243–1261.
- Kaupe, B., H. Brandt, E. M. Prinzenberg, and G. Erhardt. 2007. Joint analysis of the influence of *CYP11B1* and *DGAT1* genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *J. Anim. Sci.* 85:11–21.
- Komisarek, J., J. Szyda, A. Michalak, and Z. Dorynek. 2005. Impact of leptin gene polymorphisms on breeding value for milk production traits in cattle. *J. Anim. Feed Sci.* 14:491–500.
- Konfortov, B. A., V. E. Licence, and J. R. Miller. 1999. Re-sequencing of DNA from a diverse panel of cattle reveals a high level of polymorphism in both intron and exon. *Mamm. Genome* 10:1142–1145.
- Kuehn, C., C. Edel, R. Weikard, and G. Thaller. 2007. Dominance and parent-of-origin effects of coding and non-coding alleles at the acylCoA-diacylglycerol-acyltransferase (*DGAT1*) gene on milk production traits in German Holstein cows. *BMC Genet.* 8:62.
- Liefers, S. C., R. F. Veerkamp, M. F. te Pas, C. Delavaud, Y. Chilliard, and T. van der Lende. 2004. A missense mutation in the bovine leptin receptor gene is associated with leptin concentrations during late pregnancy. *Anim. Genet.* 35:138–141.
- Madeja, Z., T. Adamowicz, A. Chmurzynska, T. Jankowski, J. Melonek, M. Switonski, and T. Strabel. 2004. Effect of leptin gene polymorphisms on breeding value for milk production traits. *J. Dairy Sci.* 87:3925–3927.
- Näslund, J., W. F. Fikse, G. R. Pielberg, and A. Lundén. 2008. Frequency and effect of the bovine acyl-CoA:diacylglycerol acyltransferase 1 (*DGAT1*) K232A polymorphism in Swedish dairy cattle. *J. Dairy Sci.* 91:2127–2134.
- Sanders, K., J. Bennewitz, N. Reinsch, G. Thaller, E.-M. Prinzenberg, C. Kühn, and E. Kalm. 2006. Characterization of the *DGAT1*

- mutations and the CSN1S1 promoter in the German Angeln dairy cattle population. *J. Dairy Sci.* 89:3164–3174.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461–464.
- Seefried, F. R. 2008. Genomic characterisation and polymorphism analysis of candidate genes for milk production traits and association studies in three cattle breeds. PhD Thesis. Fakultät Wissenschaftszentrum Weihenstephan, Technische Universität München, Germany.
- Spelman, R. J., C. A. Ford, P. McElhinney, G. C. Gregory, and R. G. Snell. 2002. Characterization of the DGAT1 gene in the New Zealand dairy population. *J. Dairy Sci.* 85:3514–3517.
- Suzuki, R., K. Tobe, M. Aoyama, K. Sakamoto, M. Ohsugi, N. Kamei, S. Nemoto, A. Inoue, Y. Ito, S. Uchida, K. Hara, T. Yamauchi, N. Kubota, Y. Terauchi, and T. Kadowaki. 2005. Expression of DGAT2 in white adipose tissue is regulated by central leptin action. *J. Biol. Chem.* 280:3331–3337.
- Taylor, C., M. Everest, and C. Smith. 1996. Restriction fragment length polymorphism in amplification products of the bovine butyrophilin gene: Assignment of bovine butyrophilin to bovine chromosome 23. *Anim. Genet.* 27:183–185.
- Thaller, G., W. Krämer, A. Winter, B. Kaupe, G. Erhardt, and R. Fries. 2003. Effects of DGAT1 variants on milk production traits in German cattle breeds. *J. Anim. Sci.* 81:1911–1918.
- Winter, A., W. Krämer, F. A. O. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G. Thaller, and E. Fries. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk fat content. *Proc. Natl. Acad. Sci. USA* 99:9300–9305.
- Zeng, Z. B., T. Wang, and W. Zou. 2005. Modelling quantitative trait loci and interpretation of models. *Genetics* 169:1711–1725.