Genetica (2005) 125:103–113
DOI 10.1007/s10709-005-5255-1

Evaluation of candidate gene effects for beef backfat via Bayesian model selection★

Xiao-Lin Wu[1], Michael D. MacNeil[2], Sachinadan De[1,3], Qian-Jun Xiao[1], Jennifer J. Michal[1], Charles T. Gaskins[1], Jerry J. Reeves[1], Jan R. Busboom[1], Raymond W. Wright Jr.[1] & Zhihua Jiang[1,*]

[1]*Department of Animal Sciences, Washington State University, Pullman, WA 99164-6351 USA;* [2]*USDA-ARS, Fort Keogh Livestock and Range Research Laboratory, Miles City, MT 59301 USA;* [3]*Animal Biotechnology Center, National Dairy Research Institute, 132001 Karnal, Haryana India; *Author for correspondence (Phone: +509-335-8761; Fax: +509-335-4246; E-mail: jiangz@wsu.edu)*

Received 8 January 2005 Accepted 10 April 2005

Key words: backfat, Bayesian analysis, beef, candidate genes, model selection

Abstract

Candidate gene approaches provide tools for exploring and localizing causative genes affecting quantitative traits and the underlying variation may be better understood by determining the relative magnitudes of effects of their polymorphisms. Diacyglycerol *O*-acyltransferase 1 (*DGAT1*), fatty acid binding protein (heart) 3 (*FABP3*), growth hormone 1 (*GH1*), leptin (*LEP*) and thyroglobulin (*TG*) have been previously identified as genes contributing to genetic control of subcutaneous fat thickness (SFT) in beef cattle. In the present research, Bayesian model selection was used to evaluate effects of these five candidate genes by comparing competing non-nested models and treating candidate gene effects as either random or fixed. The analyses were implemented in SAS to simplify the programming and computation. Phenotypic data were gathered from a F_2 population of Wagyu × Limousin cattle. The five candidate genes had significant but varied effects on SFT in this population. Bayesian model selection identified the *DGAT1* model as the one with the greatest model probability, whether candidate gene effects were considered random or fixed, and *DGAT1* had the greatest additive effect on SFT. The SAS codes developed in the study are freely available and can be downloaded at: http://www.ansci.wsu.edu/programs/.

Abbreviations: BIC – Bayesian Information Criterion; *DGAT1* – diacyglycerol *O*-acyltransferase 1; *FABP3* – fatty acid binding protein (heart) 3; *GH1* – growth hormone 1; *LEP* – leptin; MLE – maximum likelihood estimates; QTL – quantitative trait loci; REML – residual maximum likelihood; SFT – subcutaneous fat thickness; *TG* – thyroglobulin.

Introduction

Candidate gene approaches facilitate discovering and localizing causative genes for quantitative traits (Campbell, Nonneman & Rohrer, 2003). A candidate gene can be identified as the one with biological actions involved in the development or physiology of the trait of interest (functional

★Mention of a proprietary product does not constitute a guarantee or warranty of the product by USDA, Montana AES, or the authors and does not imply its approval to the exclusion of other products that may also be suitable.

candidate gene) or chosen from the neighborhood of previous identified QTLs (positional candidate gene). Polymorphisms within selected candidate genes can be tested for association with variation in the quantitative trait so as to better understand their effects. Advantages of the candidate gene analysis include its relative robustness to genetic heterogeneity and the ability to detect small QTL effect sizes (Craddock, Dave & Greening, 2001).

Many candidate genes have been proposed as affecting subcutaneous fat thickness (SFT) in beef cattle, such as diacylglycerol *O*-acyltransferase 1 (*DGAT1*) gene (Thaller et al., 2003; Grisart et al., 2004), fatty acid binding protein (heart) 3 (*FABP3*) gene (Roy et al., 2003), growth hormone 1 (*GH1*) gene (Taylor et al., 1998), leptin (*LEP*) gene (Buchanan et al., 2002) and thyroglobulin (*TG*) gene (Barendse, 1999). However, relative magnitude of effects and thus importance of each of these genes in controlling phenotypic variation remains unknown. From the view of statistical model selection, solving this problem requires comparing competing models that may not be nested. Classic tests of hypotheses, however, have difficulty handling this situation since they require alternative models to be nested (Bollen, 1989).

For example, following the idea of classic hypothesis testing, one would first specify an initial model that includes a set of candidate genes. Next, a sequence of tests could be used based on *P*-values to decide whether the model should be simplified or expanded by reducing or increasing the number of candidate genes. The *P*-values thus obtained are limited to comparison of the two nested models and not necessarily comparable otherwise. In contrast, Bayesian model selection offers a general way for evaluating many competing models, whether they are nested or non-nested. In view of Bayesian model selection, two competing models (candidate genes) can be compared in terms of their posterior probabilities and this reasoning extends naturally to the comparison of more than two models (Congdon, 2001).

The objectives of the present research were to: (1) implement Bayesian model selection procedures using SAS® in order to simplify the programming and enhance portability of the procedures, and (2) evaluate the effects of five previously identified candidate genes on SFT.

Materials and methods

Animals and the quantitative trait

Animals used in this study were F_2 progeny derived from inter se mating of F_1 Wagyu $\times$ Limousin sires and dams. The F_1 generation was produced by Washington State University and transferred to the USDA-ARS Fort Keogh Livestock and Range Research Laboratory at Miles City, MT in the autumn of 1998. All F_2 progeny were produced at Miles City. Calves were weaned at an approximate average age of 6 months. After weaning, the calves were fed a growing ration containing approximately 12% CP and 1.0 Mcal NE_g/Kg. Approximately 4 months before the first scheduled harvest, the ration was reformulated to contain approximately 12% CP and 1.3 Mcal NE_g/Kg.

A serial harvest protocol was implemented whereby 8–10 head of cattle were harvested weekly with steers and heifers being harvested in alternate weeks. Calves born in 2000 (n = 71) were harvested beginning 8 October 2001, calves born in 2001 (n = 90) were harvested beginning 23 September 2002, and calves born in 2002 (n = 109) were harvested beginning 11 August 2003. Harvest was at a local commercial abattoir. Subcutaneous fat thickness was measured at the 12–13th rib interface perpendicular to the outside surface at a point three-fourths the length of the longissimus muscle from its chine bone end. Inaccurate measurement of subcutaneous fat thickness due to hide removal and unavailability of DNA for analysis resulted in 246 observations being available for this study. The phenotypic data were adjusted for effects of year, gender, and age at harvest (linear) before assessing the effects of the candidate genes.

Genotyping of candidate genes

Genomic DNA was isolated from white bloods cells of each animal. The primer sequences designed for genotyping these candidate genes were: *DGAT1*, forward 5′-TGGGCTCCGTGCTGGCCCTGATGGTCTA-3′ and reverse 5′-TTGAGCTCGTAGCACAGGGTGGGGGCGA-3′; *FABP3*, forward 5′-GTGAGTTGAGGAAGGGGCTGTG-3′ and reverse 5′-TAGGTCTCCACCTCTTGTCCTTCAG-3′; *GH1*, forward 5′-TGGGGT

GGGGAGGGTTCCGAATAAGGCGG-3′ and reverse 5′-TGAGGAACTGCAGGGGCCCAAG CCACGA-3′ and *TG*, forward 5′-GGGGAT-GACTACGAGTATGACTG-3′ and reverse 5′-GTGAAAATCTTGTGGAGGCTGTA-3′. Four PCR primers were designed to genotype a C/T transition in exon 2 of the *LEP* gene using tetra primer ARMS-PCR (tetra primer amplification refractory mutation system based PCR). The two outer primers were 5′-GACGATGTGCCACGTG TGGTTTCTTCTGT-3′ and 5′-CGGTTCTACCT CGTCTCCCAGTCCCTCC-3′, while the two inner primers were: 5′-TGTCTTACGTGGAGGC TGTGCCCAGCT-3′ for the T allele and 5′-AG-GGTTTTGGTGTCATCCTGGACCTTTCG-3′ for the C allele. The amplification reaction was carried out in a final volume of 10 μl, which contained 5 pmol of each primer, ~50 ng genomic DNA, 200 nM dNTPs, 2.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris–HCl, 0.1% Triton X-100 and 0.5 U of AmpliTaq Gold polymerase (Applied Biosystems, Foster City, CA). After pre-denaturation at 95 °C for 10 min, 30 amplification cycles were performed: denaturation at 94 °C for 30 s, annealing at 63 °C for 30 s and extension at 72 °C for 30 s. The reaction ended with a 5-min post-extension at 72 °C. The polymorphisms in the *TG*, *DGAT1*, *GH1*, and *FABP3* genes were identified after digestions with restriction enzymes *Bst*YI, *Crf*I, *Msp*I, and *Aci*I, respectively. PCR products or PCR-digested products were analyzed using 1.6% agarose gels, stained with ethidium bromide and photographed.

Statistical analysis

Statistical model

Consider a total of n progeny in f unrelated populations. Let y_j be a phenotypic value of individual j, which is affected by a QTL. Assuming that a candidate gene being evaluated is completely linked with the QTL (i.e. the candidate locus is the QTL), the phenotypes can be modeled as below

$$y_j = \mu + \chi_{1j}\, a + \chi_{2j}\, d + u_j + \varepsilon_j \tag{1}$$

where μ denotes the fixed population mean, a and d are additive effect and dominance effect, respectively, of the QTL, χ_{ij} and χ_{2j} are dummy variables that relate to y_j to a and d, respectively, u_j is the residual additive genetic effect that is not explained by the QTL being evaluated, and ϵ_j is a vector of residual errors. In matrix notation, the above equation becomes

$$y = 1\mu + X\beta + Zu + e \tag{2}$$

where β is a vector that includes the additive (a) and dominance (d) effects of the candidate gene, and X and Z are incidence matrices that relate phenotypes of individual animals in y to model effects in β and u, respectively. Further, if we express y_j as derivation from the overall mean, equation is simplified as

$$y = X\beta + Zu + e \tag{3}$$

This is a mixed model if we treat candidate gene effects as fixed variables. Assuming that u and e are independent, the covariance matrix of y is

$$V = Var(y) = ZAZ'\sigma_u^2 + I\sigma_\varepsilon^2 \tag{4}$$

where A is the numerator relationship matrix. Alternatively, if candidate gene effects are treated as random variables, equation (3) is then a random model where, under the assumption of independence of a, d, and u, the covariance matrix of y is

$$\begin{aligned} V &= Var(y) \\ &= \begin{bmatrix} X_a' X_a & 0 \\ 0 & X_d' X_d \end{bmatrix} \begin{bmatrix} \sigma_a^2 \\ \sigma_d^2 \end{bmatrix} + ZAZ'\sigma_u^2 + I\sigma_\varepsilon^2 \end{aligned} \tag{5}$$

Bayesian model selection

Competing models are compared based on the Bayesian Information Criterion (BIC), which is a large sample approximation to twice the logarithm of the Bayes factor (Schwarz, 1978). Using BIC facilitates rapid computation of approximated Bayes factor from the output of some commonly used statistical software packages. The Bayesian information criteria that favors candidate gene 1 over candidate gene 2 is given as

$$\begin{aligned} BIC_{12} &= -2\{\log P(y|\hat{\theta}_1, M_1) - \log P(y|\hat{\theta}_2, M_2)\} \\ &\quad + (p_1 - p_2)\log(n) \\ &= -2\{\log P(y|\hat{\theta}_1, M_1) - \log P(y|\hat{\theta}_2, M_2)\} \end{aligned} \tag{6}$$

where $P(y|\hat{\theta}_1, M_1)$ and $P(y|\hat{\theta}_2, M_2)$ are the maximized likelihood under models M_1 and M_2, respectively, $\hat{\theta}_1$ represents the MLE (maximum likelihood estimates) under Model M_1 whose

dimension is p_1, $\hat{\theta}_2$ represents the MLE of Model 2 parameter whose dimension is p_2, and n is the sample size. Note that the last term, $(p_1–p_2)\log(n)$, cancels out since the competing models are of the same dimension (i.e., $p_1 = p_2$) in the present analyses.

More intuitively, two competing models can be evaluated in terms of their posterior probability

$$p(M_i|y) = \frac{P(y|M_i)}{P(y|M_1) + p(y|M_2)}, \quad i = 1, 2 \qquad (7)$$

This reasoning extends very naturally to the situation when $k > 2$ candidate genes were compared such that

$$p(M_i|y) = \frac{P(y|M_i)}{P(y|M_1) + P(y|M_2) + \cdots + P(y|M_k)}, \quad \mathrm{i} = 1, 2, \ldots, k \qquad (8)$$

REML estimation of candidate gene effects

Consider the mixed model where candidate gene effects are treated as fixed variables. In the residual maximum likelihood (REML) analysis, the following objective function associated with REML (i.e. residual log likelihood) was constructed based on model (3) and maximized over all unknown parameters (actually, the PROC MIXED minimizes –2 times the following function using a ridge-stabilized Newton–Raphson algorithm).

$$l_R = -\frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| - \frac{1}{2}r'V^{-1}r - \frac{n-p}{2}\log(2\pi) \qquad (9)$$

where $r = y–X(X'V^{-1}X)^{-}X'V^{-1}y$ and p is the rank of X. Model effects are estimated following standard mixed model theory (Henderson, 1984). A general t-statistic is constructed for inferences concerning the fixed and random parameters in the model. REML estimation of random candidate gene effects (random models) using the MIXED procedure is a special case of mixed model approach without fixed effects (or the overall mean being the only fixed effect).

Bayesian estimation of candidate gene effects

The structure of the mixed model lends itself to a two-step approach to sampling posteriors (Wolfinger & Kass, 2000). Again, consider the mixed model, and the random model is treated as a special case of mixed model. Let ω be the vector of variance components in the model, the joint posterior density of (β, u, ω) is written as

$$p(\beta, u, \omega|y) = p(\beta, u|\omega, y)p(\omega|y) \qquad (10)$$

The two distributions on the right-hand side can be considered separately. Sampling from the two distributions comprises the two primary steps of generating posterior samples in our Bayesian analysis.

Following Bayes theorem, the marginal posterior for unknown variances can be obtained as

$$p(\omega|y) \propto \pi(\omega)p(y|\omega) \qquad (11)$$

where $p(y|\omega)$ is the integrated likelihood function and $\pi(\omega)$ is the prior for ω. In the analysis, a flat prior was assumed for β and Jeffrey's prior for ω. The conditional posterior for β and u is relatively easy to obtain following standard Gaussian mixed model theory, which takes a form of multivariate normal density (Henderson, 1984). In the Bayesian analysis, posterior sampling is conducted in the following two steps for predefined cycles of updates: (1) Sample ω_* from $p(\omega|y)$ using the independence chain algorithm; and (2) Given ω_* from step 1, β and u are sampled from the multivariate normal distribution $p(\beta, u|\omega, y)$.

The independence chain algorithm (Tierney, 1994) used in the first step requires a fixed base sampling density $g(\omega|y)$, from which proposals are drawn. The base density $g(\omega|y)$ is defined through using products of inverted-gamma density. The chain starts with a pseudo-random draw from $g(\omega|y)$. Then, supposing that the chain is currently at state ω_t, the algorithm generates a candidate ω_* from $g(\omega|y)$ and accepts it with the following probability

$$\min\left[\frac{q(\omega_t|\omega_*)p(\omega_*)}{q(\omega_*|\omega_t)p(\omega_t)}, 1\right] \qquad (12)$$

Otherwise, a copy of ω_t is added to the chain.

Implementing Bayesian analysis in the SAS system

Bayesian analysis was implemented using PROC MIXED (SAS Institute Inc., Cary, NC). PROC MIXED used an independence chain algorithm to generate the posterior samples for the variance components. The IC algorithm works by generat-

ing a pseudo-random proposal from a convenient base distribution, chosen to be as close as possible to the posterior. The proposal is then retained in the sample with probability proportional to the ratio of weights constructed by taking the ratio of the true posterior to the base density. Posterior samples were also generated for fixed effects given estimated variance components with the solution option specified in the model statement. For each parameter in the model, a single chain of 20,000 updates was generated, which were later thinned to 2000 posterior samples. These samples were used for making inference of marginal distributions of model parameters.

Given that the additive relationship matrix A is provided from an input file, the SAS MIXED procedure performed well with the REML analysis, but it had difficulty sampling the posteriors for the residual genetic variance. To overcome this limitation, for example with fixed candidate gene effects defined in the model, we sampled model using the following multivariate normal density.

$$p(\beta,u|\varpi,y) \propto MVN \times \left(\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+\lambda_u^{-1} \end{bmatrix}^{-} \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+\lambda_u^{-1} \end{bmatrix} \right) \quad (13)$$

where $\lambda_u = \sigma_u^2/\sigma_e^2$. This simplification was not necessary with Bayesian model selection, since Bayesian information content was not calculated using posterior data.

Results and discussion

Bayesian evaluation of candidate genes on SFT

Information on mutation, sample size, genotype frequencies and allele frequencies for *DGAT1*, *FABP3*, *GH1*, *LEP* and *TG* genes was listed in Table 1. A single candidate gene was included in the model in a separate analysis. As different candidate genes were treated as different QTL with various effects and directly evaluated effects of these candidate genes on the quantitative trait, the present single candidate gene analysis was straightforward. Alternatively, it is possible to evaluate the five candidate genes simultaneously with a consequent increase in the number of parameters to be estimated, computational complexity, and cost. In particular, simultaneous evaluation of multiple candidate genes may give rise to the need to consider many interactions in the fixed-effect model or many covariances in the random-effect model. Thus, the present approach

Table 1. Loci, mutations, sample size and genotypes and allele frequencies for polymorphisms in five candidate genes

Locus	Mutation	Sample size	Genotype (frequency)	Alleles (frequency)
TG	C/T	242	CC: 295 + 178 + 75 bp (38.84%)	C (60.95%)
			CT: 473 + 295 + 178 + 75 bp (44.21%)	T (39.05%)
			TT: 473 + 75 bp (16.94%)	
LEP	C/T (exon 2)	246	CC: 239 + 164 bp (44.71%)	C (67.89%)
			CT: 239 + 164 + 131 bp (46.34%)	T (32.11%)
			TT: 239 + 131 bp (8.94%)	
DGAT1	C/A	246	AA: 405 bp (34.55%)	A (58.33%)
			AC: 175 + 230 + 405bp (47.56%)	C (41.67%)
			CC: 175 + 230 bp (17.89%)	
GH1	*Msp*I (intron 3)	243	++: 223 + 109 + 97 + 63 bp (81.07%)	+(90.12%)
			+−: 332 + 223 + 109 + 97 + 63 bp (18.11%)	−(9.88%)
			− −: 332 + 97 + 63bp (0.82%)	
FABP3	A/G	243	AA: 438 bp (6.58%)	A (24.69%)
			AG: 438 + 299 + 139 bp (36.21%)	G (75.31%)
			GG: 299 + 139 bp (57.20%)	

Table 2. Fixed candidate gene additive and dominance effects and residual variance estimated using restricted maximum likelihood

	TG	*LEP*	*DGAT1*	*GH1*	*FABP3*
Additive effect					
Mean	−0.0016	0.0190	0.0318*	−0.0232	−0.0155
SE	0.0211	0.0187	0.0148	0.0250	0.0164
Dominance effect					
Mean	0.0063	0.0184	0.0019	−0.0307	0.0195
SE	0.0148	0.0121	0.0106	0.0227	0.0130
Residual variance					
σ_e^2	0.0243	0.0243	0.0243	0.0243	0.0243

*$P < 0.05$.

might be viewed as individual validations of five previously identified candidate genes using one test population. From these validation exercises, comparisons among the candidate gene effects have been extracted.

Two methods were used to estimate effects of these candidate genes: REML and Bayesian estimation. REML estimation (Table 2) identified *DGAT1* as having significant additive effect on **SFT** ($P < 0.05$). A large additive effect of *GH1* was also observed but it was not significant due to a large standard error. None of dominance effect estimates were found significant ($P > 0.05$). In contrast to REML estimation, Bayesian analysis provided more detailed additive and dominance effects through their posterior distributions (Figure 1). For example, the posterior mean of the additive effect of DGAT1 on SFT was 0.0319, and the 5, 25, 50, 75, and 95% tiles for its additive effect were 0.0070, 0.0216, 0.0320, 0.0421, and 0.0572, respectively. Similarly, the posterior means for its dominance effect was 0.0017. The 5, 25, 50, 75, and 95% tiles for its additive effect were −0.0160, −0.0560, 0.0016, 0.0088, and 0.0195, respectively. However, both methods identified *DGAT1* as having the greatest additive effect (REML: 0.0318 ± 0.0148; Bayesian: 0.0319 ± 0.0150) and *GH1* as having the greatest dominance effect (REML: -0.0307 ± 0.0227; Bayesian: -0.0219 ± 0.0230) on **SFT**. The student *t*-test indicated that the REML estimate of *DGAT1* additive effect was significantly greater than zero ($P < 0.05$). Similarly, Bayesian analysis showed that 98.29% of posterior samples of *DGAT1* additive effect were positive values (Figure 1a). Thus, both results indicate that *DGAT1* significantly affected **SFT** in this beef cattle population. Using either the REML or Bayesian analysis, estimated *GH1* dominance effects were also similar and not significant in view of either the student *t*-test (REML: $P = 0.4743$) or posterior distribution (Bayesian: 49.95% of the posteriors of the *GH1* dominance effect were positive while 50.05% of them were negative, Figure 1b).

Bayesian estimation of fixed candidate gene effects was in good agreement with the REML estimation, but considerable differences were found in estimated random candidate gene effects between these two methods (Figure 2). Noticeably, the REML analysis had difficulty estimating candidate gene variance when only two alleles were involved and it tended to give zero estimates whenever they should not be. In contrast, Bayesian estimation dealt with variance estimation in the bi-allelic systems more naturally, though not a complete success. Regarding residual variance, both methods gave similar and consistent estimation (Figure 2).

Bayesian model selection consistently indicated that an allelic substitution at the DGAT1 locus had the greatest influence on the thickness of the subcutaneous fat of animals harvested on an age-constant basis from this Wagyu–Limousin population in the five candidate gene loci we examined, based on either fixed candidate gene effect models or random candidate gene effect models (Figure 3). Compared to the null model without any candidate gene, models that contained each of the five genes had significantly greater model probability (i.e. BICs varied from –177.1 to

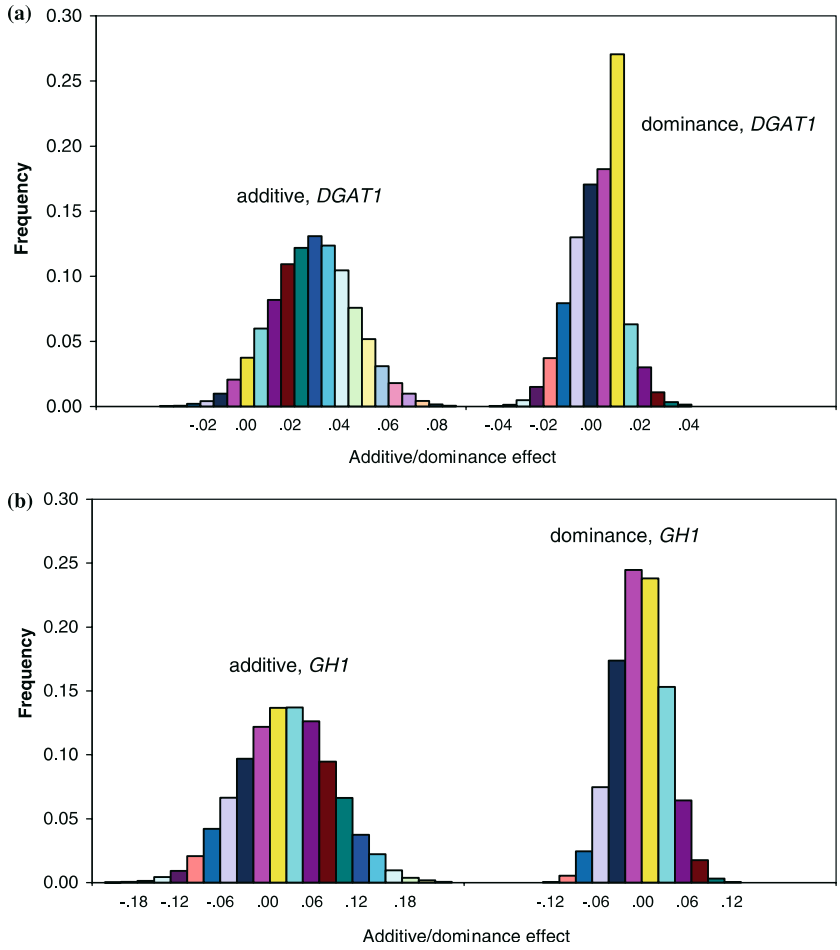

Figure 1. Posterior distributions of: (a) *DGAT1* and (b) *GH1* candidate gene effects on age-constant subcutaneous fat thickness in Wagyu × Limousin beef cattle. Posterior samples were collected at every 10th of the 20,000 updates based on models with fixed candidate gene effects.

–182.9). Thus, all five genes contributed significantly to variation to **SFT** in this beef cattle population, but with varied importance. When treating candidate gene effects as fixed variables, Bayesian model selection identified the *DGAT1* model as having the highest model probability (54.08%), followed by the *GH1* model (21.99%). In contrast, the model probabilities for the other three candidate genes were relatively low (from 2.56 to 11.47%). Similarly, Bayesian model selection with random candidate gene effects also identified the DGAT1 model as having the greatest model probability (47.9%).

The *DGAT1* gene encodes diacylglycerol *O*-acyltransferase 1, which is a microsomal enzyme that plays a central role in the metabolism of cellular diacylglycerol lipids and catalyzes the only committed step in triacylglycerol synthesis by using diacylglycerol and fatty acyl CoA as substrates. A lysine/alanine polymorphism of the *DGAT1* gene was previously reported to be associated with milk fat content (Winter et al., 2002; Grisart et al., 2004). *DGAT1* was suggested as a functional and positional candidate gene for milk fat content with evidence coming from: (1) the phenotype of *DGAT1*-deficient mice (Smith et al., 2000) and (2) its position close to a milk fat QTL on bovine chromosome 14 (Riquet et al., 1999). It was hypothesized that a lysine residing at position 232 of the *DGAT1* protein could confer more efficient binding of acyl-coenzyme A than an alanine residing at that position (Winter et al., 2002). Thaller et al. (2003) further suggested *DGAT1* as a new positional and functional

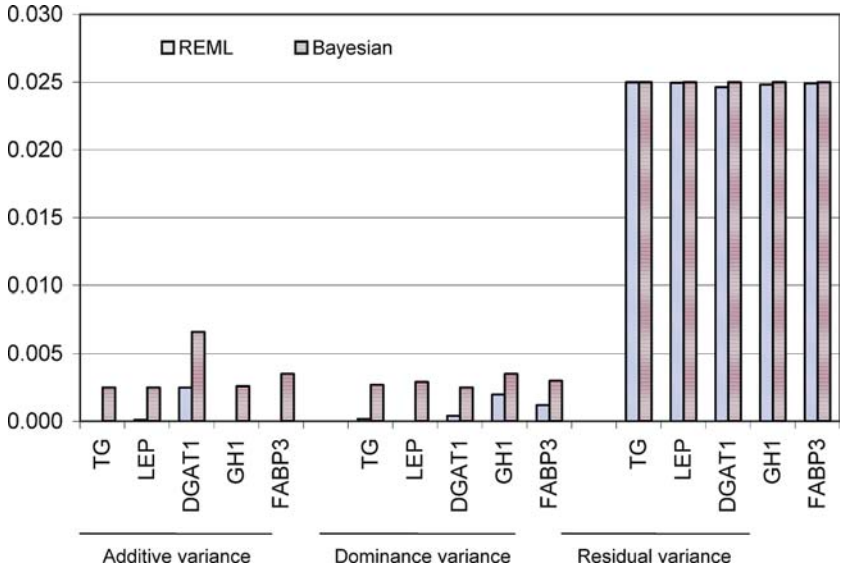

Figure 2. Estimates of additive, dominance, and residual variance components from restricted maximum likelihood and Bayesian analyses of age-constant subcutaneous fat thickness in Wagyu × Limousin beef cattle.

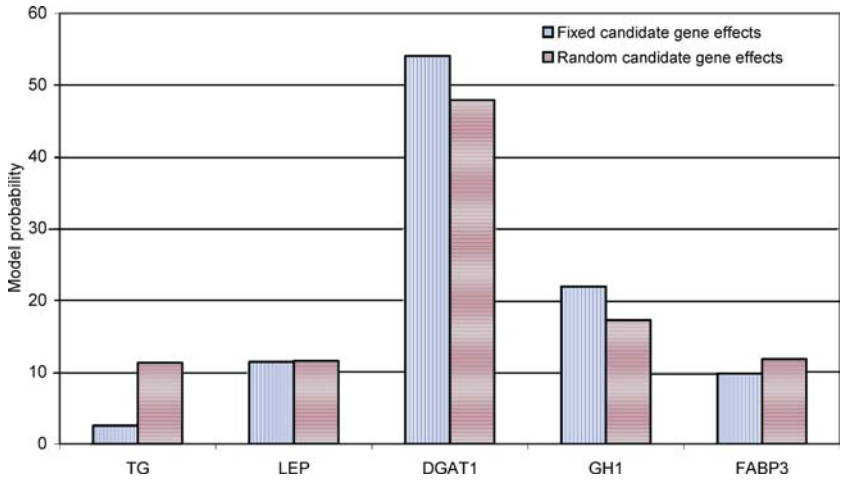

Figure 3. Model probabilities for *TG*, *LEP*, *DGAT1*, *GH1* and *FABP3* loci affecting age-constant subcutaneous fat thickness in Wagyu × Limousin beef cattle.

candidate gene for intramuscular fat deposition in cattle because: (1) its product is directly involved in triglyceride synthesis (Cases et al., 1998), (2) expressed sequence tags analysis has found that it is expressed in adipose tissue (Fries and Winter, 2002), in addition to the bovine mammary gland, and (3) radiation hybrid mapping has placed *DGAT1* on chromosome 14 (Womack et al., 1997), proximal to CSSM66, a microsatellite marker that was associated with a lipid QTL (Barendse, 1999). In humans, Smith et al. (2000) has suggested that the selective inhibition of *DGAT1*-mediated triglyceride synthesis may be useful for treating obesity.

Bayesian analysis in the SAS system

The SAS system has been widely used in analyzing QTL and candidate genes in the past decades, yet it has not been known to support much

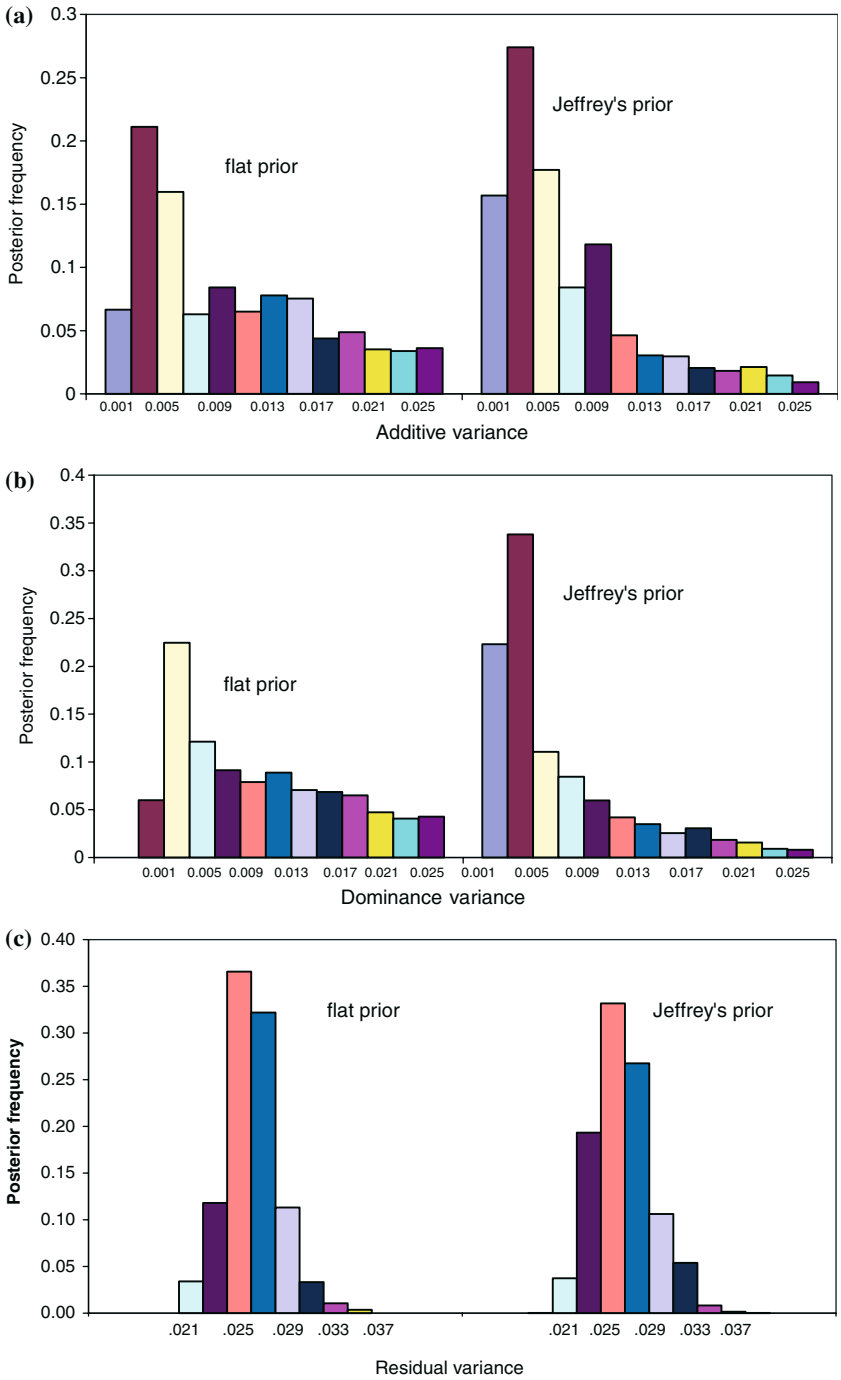

Figure 4. Comparison of Jeffrey's and flat priors on estimation of: (a) additive variance, (b) dominance variance, and (c) residual variance components. Posterior samples were collected at every 10th of the 20,000 updates based on model with random candidate gene effects.

Bayesian analysis. Bayesian analyses have been incorporated into the PROC MIXED procedure with the availability of the PRIOR statement in the SAS system of version 7 or later. It currently operates with variance component models and can estimate their marginal posterior density.

PROC MIXED allows for various strategies to generate posterior samples for variance components, which include: independence chain (IC), importance sampling (IS), rejection sampling (RS), and random walk chain (RWC). An independent chain algorithm was used to generate the posterior sample in the present analysis. As a particular MCMC alternative to Gibbs sampling, the IC algorithm has many features making it different from the latter, such as sampling variance components as a block, and using the same proposal density through the entire algorithm so that it requires no burn-in updates. Practically, the IC algorithm is more efficient to handle unbalanced, nonconjugate cases than the Gibbs sampling (Wolfinger & Kass, 2000).

Under the Bayesian framework, the type of model effects, whether fixed or random, depends on how we treat the prior distributions (Yi & Xu, 2000). In this research, alternative analyses of the data consistently supported each other. However, estimation of variance components was not entirely satisfactory in these bi-allelic systems. Xu (1998) found that treating QTL as random variables is justifiable with multiple alleles than with just two alleles. Treating bi-allelic gene effects as random in the model may be accompanied by concern about bias in estimated genetic variances. As shown in a previous research, estimated genetic variances based on random-QTL models vary dramatically depending on sampling strategies, and sampling more alleles (families) would greatly improve estimation of genetic variances (Wu & Jannink, 2004). We also observed a slight difference in the Markov chain behavior when treating candidate gene effects as either fixed or random variables. Chains for random candidate gene effects were less movable than for fixed candidate gene effects, possibly due to each candidate gene having only two alleles. The bi-allelic system would cause difficulty for the chain to move to a new state. Nevertheless, the sampling acceptance probability for the five random candidate gene effects ranged from 0.13 to 0.29, which was still within the range of the suggested acceptance probability (Roberts, 1995; Jannink & Wu, 2003).

The use of Jeffrey's prior influenced posterior distributions of both additive and dominance variances relative to use of flat priors (Figures 3 and 4), when candidate gene effects were treated as random variables. Posteriors of the two variance components obtained using Jeffrey's prior were more heavily distributed toward smaller values. This situation is typical in small data sets where the prior dominates the posterior. In contrast, estimates of residual variance were insensitive to method of analysis, fixed versus random candidate gene effects, and choice of prior.

Acknowledgements

Agricultural Research Center, Washington State University and Blue Rock Inc. supported this project. We thank Dr Shizhong Xu for helpful discussion regarding the statistical analysis. Thanks also go to the anonymous reviewers and the editor for their valuable comments on the manuscripts.

References

Barendse, W.J., 1999. Assessing lipid metabolism. Patent International Publication number: WO 99/23248. World International Property Organization.

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.

Buchanan, F.C., C.J. Fitzsimmons, A.G. Van Kessel, T.D. Thue & D.C. Winkelman, 2002. Association of a missense mutation in the bovine leptin gene with carcass fat content and leptin mRNA levels. Genet. Sel. Evol. 34: 105–116.

Cases, S., S.L. Smith, Y.W. Zheng, H.M. Myers & S.R. Lear, 1998. Identification of a gene encoding an acyl CoA: diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proc. Natl. Acad. Sci. USA 95: 13018–13023.

Campbell, E.M., D. Nonneman & G.A. Rohrer, 2003. Fine mapping a quantitative trait locus affecting ovulation rate in swine on chromosome 8. J. Anim. Sci. 81: 1706–1714.

Congdon, P., 2001. Bayesian Statistical Modeling. John Wiley & Sons, New York.

Craddock, N., S. Dave & J. Greening, 2001. Association studies of bipolar disorder. Bipolar Disord. 3: 284–298.

Fries, R. & A. Winter, 2002. Method of testing a mammal for its predisposition for fat content of milk and/or its predisposition for meat marbling. Patent, International Application Number: PCT/EP 02/07520. World International Property Organization.

Grisart, B., F. Farnir, L. Karim, N. Cambisano & J.J. Kim, 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. Application of Linear Models in Animal Breeding. University of Guelph, Guelph, Ontario, Canada.

Jannink, J.-L. & X.-L. Wu, 2003. Estimating allelic number and identity in state of QTLs in interconnected families. Genet. Res. 81: 133–144.

Riquet, J., W. Coppieters, N. Cambisano, J.J. Arranz & P. Berzi, 1999. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. Proc. Natl. Acad. Sci. USA 96: 9252–9257.

Roberts, G.O., 1995. Markov chain concepts related to sampling algorithm, pp. 45–48 in Markov Chain Monte Carlo in Practice, edited by W.R. Gilks, S. Richardson & D.J. Spiegelhalter. Chapman & Hall, London.

Roy, R., J.H. Calvo, H. Hayes, C. Rodellar & A. Eggen, 2003. Fine mapping of the bovine heart fatty acid-binding protein gene (FABP3) to BTA2q45 by fluorescence *in situ* hybridization and radiation hybrid mapping. Anim. Genet. 34: 466–467.

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

Smith, S.J., S. Cases, D.R. Jensen, H.C. Chen & E. Sande et al., 2000. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking DGAT. Nat. Genet. 25: 87–90.

Taylor, J.F., L.L. Coutinho, K.L. Herring, D.S. Gallagher & R.A. Brenneman, 1998. Candidate gene analysis of GH1 for effects on growth and carcass composition of cattle. Anim. Genet. 29: 194–201.

Thaller, G., C. Kuhn, A. Winter, G. Ewald & O. Bellmann, 2003. *DGAT1*, a new positional and functional candidate gene for intramuscular fat deposition in cattle. Anim. Genet. 34: 354–357.

Tierney, L., 1994. Markov chains for exploring posterior distribution (with discussion). Ann. Stat. 22: 1701–1762.

Winter, A., W. Kramer, F.A. Werner, S. Kollers & S. Kata, 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.

Wolfinger, R.D. & R.E. Kass, 2000. Non-conjugate Bayesian analysis of variance component models. Biometrics 56: 768–774.

Womack, J.E., J.S. Johnson, E.K. Owens, C.E. Rexroad III & J. Schlapfer, 1997. A whole-genome radiation hybrid panel for bovine gene mapping. Mamm. Genome 8: 854–856.

Wu, X.-L. & J.-L. Jannink, 2004. Optimal sampling of a population to determine QTL location, variance, and allelic number. Theor. Appl. Genet. 108: 1434–1442.

Xu, S., 1998. Mapping quantitative trait loci using multiple families of line crosses. Genetics 148: 517–524.

Yi, N. & S. Xu, 2000. Bayesian mapping of quantitative trait loci under the identity-by-descent-based variance component model. Genetics 156: 411–422.