Genet. Res., Camb. (2006), **88**, *pp.* 143–151. © 2006 Cambridge University Press
doi:10.1017/S0016672307008488 Printed in the United Kingdom

A whole-genome analysis using robust asymmetric distributions

LUIS VARONA[1]*, WAGDY MEKKAWY[2], DANIEL GIANOLA[3,4] AND AGUSTÍN BLASCO[2]

[1] *Area de Producció Animal, Centre UdL-IRTA, Av. Rovira Roure 191, 2519 Lleida, Spain*
[2] *Departamento de Ciencia Animal, Universidad Politécnica de Valencia, 46071 Valencia, Spain*
[3] *Departments of Animal Sciences, Dairy Science and Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI 5370, USA*
[4] *Department of Animal and Aquacultural Sciences, Agricultural University of Norway, 1432 Ås, Norway*

(*Received 25 July 2006 and in revised form 22 November 2006*)

Summary

This study is aimed at improving the analysis of data used in identifying marker-associated effects on quantitative traits, specifically to account for possible departures from a Gaussian distribution of the trait data and to allow for asymmetry of marker effects attributable to phenotypic divergence between parental lines. A Bayesian procedure for analysing marker effects at the whole-genome level is presented. The procedure adopts a skewed *t*-distribution as a prior distribution of marker effects. The model with the skewed *t*-process includes Gaussian prior distributions, skewed Gaussian prior distributions and symmetric *t*-distributions as special cases. A Markov Chain Monte Carlo algorithm for obtaining marginal posterior distributions of the unknowns is also presented. The method was applied to a dataset on three traits (live weight, carcass length and backfat depth) measured in an F_2 cross between Iberian and Landrace pigs. The distribution of marker effects was clearly asymmetric for carcass length and backfat depth, whereas it was symmetric for live weight. The *t*-distribution seems more appropriate for describing the distribution of marker effects on backfat depth.

1. Introduction

Recent development of molecular techniques has provided a massive number of molecular markers and, as a consequence, dense genetic maps are now available for a number of species. An obvious use of this molecular information is for marker-assisted selection in livestock and plant populations (Whittaker, 2001). The basic idea of marker-assisted selection is to detect and exploit linkage disequilibrium between mutations that cause genetic variations and presumably neutral molecular markers.

Over the last decade, many quantitative trait loci (QTL) mapping experiments have been performed to detect regions of the genome involved in genetic regulation of quantitative traits (Pe *et al*., 1993; Devicente *et al*., 1993; Andersson *et al*., 1994). In these studies, statistical analyses were usually carried out via a genomic scan using likelihood (Lander & Botstein, 1989) or regression techniques (Haley *et al*., 1994). With these approaches, the model of analysis allows for one or a few genome locations at a time.

Recently some authors have proposed the combined use of all available markers. For instance, Whittaker (2001) and Lange & Whittaker (2001) proposed ridge regression methods, while Gianola *et al*. (2003) and Xu (2003) developed Bayesian approaches. The procedures of Gianola *et al*. (2003) and Xu (2003) employ prior Gaussian distributions for effects associated with molecular markers. Mekkawy (2005) compared the two procedures by simulation, and showed that the method of Gianola *et al*. (2003) produced better predictions of genetic merit.

However, the assumption of a prior Gaussian distribution, common to both Gianola *et al*. (2003) and Xu (2003), for effects associated with molecular markers may be unrealistic if some of the markers are associated with major effects. Therefore, it is possible that a more robust distribution, such as a

* Corresponding author. Telephone: +34 973003441. Fax: +34 973238301. e-mail: Luis.varona@irta.es

t-distribution, could be more appropriate (Lange *et al*., 1989) as prior. Further, asymmetric distributions (Fernandez & Steel, 1998) may provide a more flexible prior for effects associated with genetic markers.

The objective of this paper is to describe a procedure based on Gianola *et al*. (2003), but with robust asymmetric prior distributions instead, and to evaluate its performance in a case involving data from a crossing experiment with Iberian and Landrace pigs.

2. Asymmetric robust distribution

Under the assumption of prior independence between markers, Gianola *et al*. (2003) assume the following Gaussian prior distribution for additive marker effects:

$$f(\mathbf{q}|\sigma_q^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left[-\frac{q_i^2}{2\sigma_q^2}\right]$$

where $\mathbf{q} = \{q_i\}$ is the vector of additive genetic marker effects, σ_q^2 is the variance of the additive genetic marker effects and n is the number of markers.

Here, we present an alternative prior distribution following Fernandez & Steel (1998), with density:

$$f(\mathbf{q}|\sigma_q^2, \gamma) = \prod_{i=1}^{n} \frac{1}{\gamma + \frac{1}{\gamma}} \left\{ f\left(\frac{q_i}{\gamma}\right) I_{[0,\infty)}(q_i) + f(q_i\gamma) I_{(-\infty,0]}(q_i) \right\} \tag{1}$$

where γ is the asymmetry parameter, and $f(x)$ is the following univariate *t*-density:

$$f(x) = \frac{\Gamma\left[\frac{1}{2}(v+1)\right]}{\sigma_q\sqrt{v\pi}\,\Gamma\left(\frac{1}{2}v\right)} \left(1 + \frac{x^2}{v\sigma_q^2}\right)^{\frac{-(v+1)}{2}} \tag{2}$$

with v representing the degrees of freedom parameter of the *t*-distribution.

3. Application

(i) *Experimental design*

Three Iberian boars from the genetically isolated Guadyerbas line (Toro *et al*., 2000) were mated with 31 non-inbred Landrace sows. Six boars and 73 sows from their offspring, generation F_1, were the parents of 577 F_2 animals. The Iberian pig is characterized by its extremely fat body composition. Landrace animals were the maternal line used at Nova Genetica S. A. experimental farm. The two parental lines differ substantially in growth, carcass and meat quality traits (Serra *et al*., 1998).

The F_2 pigs were raised under standard intensive conditions on the experimental farm Nova Genetica. Feeding was *ad libitum*, and males were not castrated. A total of 321 individuals from 58 full-sib families were recorded for carcass weight, carcass length and backfat depth. The average ($\pm$SD) age at slaughter was $175{\cdot}5 \pm 5{\cdot}5$ days. Average ($\pm$SD) carcass weight, carcass length and backfat depth were $74{\cdot}9 \pm 9{\cdot}82$ kg., $79{\cdot}26 \pm 3{\cdot}96$ cm and $28{\cdot}31 \pm 7{\cdot}90$ mm, respectively.

(ii) *Genotyping*

DNA of parent animals was extracted from blood using a saline precipitation protocol, and DNA from F_1 and F_2 pigs was extracted using a commercial protocol (Boehringer Mannheim). Animals were genotyped for 92 markers (90 microsatellites and 2 PCR-RFLP), chosen to provide highly informative input based on the index of Ron *et al*. (1995). A broad description of the markers is presented in Table 1. Markers provided a uniform coverage of the 18 autosomes. PCRs were carried out in a MJ Research Thermal Cycler. The microsatellite PCR products were analysed with Genescan software on capillary electrophoresis equipment with fluorescence detection (ABI PRIMS 310 Genetic Analyzer). Genotypes were stored in the Gemma database (Iannucelli *et al*., 1996).

(iii) *Statistical analysis*

Statistical analysis of carcass weight, carcass length and backfat depth for the F_2 population was carried out using univariate Bayesian methods. The likelihood of the data given the parameters of the model was

$$f(\mathbf{y}|\boldsymbol{\beta}, \mathbf{q}, \sigma_e^2, \mathbf{M}) \propto \frac{1}{\sigma_e^{n_d}} \exp\left[-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{M}\mathbf{q})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{M}\mathbf{q})}{2\sigma_e^2}\right]$$

where $\mathbf{y}$ is the vector of data (321 records); $\boldsymbol{\beta}$ is the vector of systematic effects (2 levels for sex, 58 levels for family, and a regression on age of slaughter); $\mathbf{q}$ is the vector of additive effects associated with the markers (92 levels); $\mathbf{X}$ is an incidence matrix relating the systematic effects to the data; $\mathbf{M}$ is an unknown incidence matrix relating the additive marker effects to the data; σ_e^2 is the residual variance; and n_d is the number of data points.

The incidence matrix $\mathbf{M}$ is an n_d (number of data points) $\times$ n_m (number of markers) random matrix, with entries equal to 1, 0 or -1, depending on whether the two alleles were of Iberian origin (1), one was Iberian and the other was a Landrace allele (0), or both were Landrace alleles (-1). The prior distribution of the matrix was determined by the probability of origin given the marker data ($f(\mathbf{M}|Mks)$), which was calculated using the algorithm

Table 1. *Marker positions* (*Pos*) *and information content* (*IC*) *at position sorted by chromosome* (*Chr*)

Chr	Marker	Pos	IC	Chr	Marker	Pos	IC	Chr	Marker	Pos	IC
1	SW1515	0·0	0·53	6	S0035	0·0	0·65	12	S0143	0·0	0·78
	CGA	30·1	0·99		SW1057	44·3	0·96		GH	31·4	0·71
	S0113	46·2	0·57		S0087	57·7	1·00		SW874	48·6	0·98
	S0155	55·0	0·78		SW316	81·2	0·90		S0106	81·7	0·79
	SW1828	85·0	0·85		S0228	96·0	0·48	13	S0219	0·0	0·57
2	IGF2	0·0	0·70		SW1881	108·7	0·82		SW935	30·9	0·54
	S0141	30·3	0·93		SW2419	145·3	0·96		SWR100	64·0	0·97
	SW240	41·8	0·98	7	S0025	0·0	0·73		SW398	81·4	0·96
	SW395	64·7	0·95		S0064	40·1	0·76		SW1056	91·2	0·41
	S0226	72·4	0·99		TNFB	68·9	0·93		SW769	121·5	0·58
	S0378	87·0	0·93		S0066	87·8	0·99	14	SW857	0·0	0·89
	SWR308	130·1	1·00		SW632	111·9	0·94		SW1125	18·8	0·92
3	SW72	0·0	1·00		S0101	137·7	0·92		SW210	42·2	0·82
	S0206	25·6	0·51		SW764	160·3	0·94		S0007	55·8	0·87
	S0216	55·1	1·00	8	SW2410	0·0	0·98		SW1557	90·8	0·43
	S0002	77·5	0·72		SW905	26·0	0·60		SW2515	114·0	0·75
	Sw349	86·0	0·99		SWR110	44·7	1·00	15	SW919	0·0	1·00
4	SW2404	0·0	0·80		S0017	66·5	0·98		SW1111	16·3	0·56
	S0301	40·8	0·85		S0225	86·1	0·83		S0149	38·3	1·00
	S0001	59·5	0·88		SW61	109·1	1·00		SW936	56·0	0·85
	SW839	72·8	1·00	9	SW983	0·0	0·99		SW1119	79·9	0·45
	DECR2	78·8	0·18		SW911	31·1	0·71	16	SW742	0·0	0·99
	S0214	95·1	1·00		SW2571	79·5	0·73		S0298	18·4	0·44
	SW445	116·8	1·00		SW2093	109·2	0·99		SW2517	35·9	1·00
	S0097	134·4	0·84		SW1349	160·9	0·76		S0061	69·4	0·37
5	SW413	0·0	0·98	10	S0038	0·0	0·74	17	SW24	0·0	1·00
	SW2425	66·1	0·50		S0070	45·5	0·82		SW1920	28·3	0·95
	S0005	81·8	1·00		SW1626	100·5	0·97		SW2431	72·2	0·51
	IGF1	113·8	0·91	11	S0385	0·0	0·87	18	SW1023	0·0	0·85
	SWR111	130·9	0·95		S0071	43·1	0·75		SW787	21·5	0·84
					SW703	72·3	1·00		S0120	35·1	1·00

described by Haley *et al.* (1994); *Mks* stands for markers.

The prior distribution of the additive marker effects was as described in the previous section in expressions (1) and (2). In addition, the prior distribution for systematic effects ($\boldsymbol{\beta}$) was:

$$f(\beta|\beta^0, \sigma_\beta) = N(\mathbf{1}\beta^0, \mathbf{I}\sigma_\beta)$$

where $\mathbf{1}\ \beta^0$ is the vector of prior means for the systematic effects, ($\mathbf{1}$ is an $n_m \times 1$ vector of ones) and σ_β is the prior standard deviation. Here, β^0 was set to 0, and σ_β was set to 10^6. Prior distributions of the variance components (σ_e^2 and σ_q^2) were set to

$$f(\sigma_e^2) = \chi^{-2}(\nu_e, s_e^2)$$

$$f(\sigma_q^2) = \chi^{-2}(\nu_q, s_q^2)$$

with $\nu_e = \nu_q = 0$, and $s_e^2 = s_q^2 = -2$, thus yielding uniform distributions. In addition, the prior distribution of the degrees of freedom of the *t*-distribution (ν) was flat between 1 and 100. Finally, the prior distribution of the asymmetry parameter (γ) was assumed to be uniform between 0 and 1 with density of 0·5, and with a density of $\frac{0{\cdot}5}{\gamma^2}$ for values over 1. This distribution assigns an equal probability (0·5) to values below and over 1, and it is invariant with respect to an arbitrary choice of the sign of the genetic effects favouring the Landrace or the Iberian line. This prior distribution is presented in Fig. 1.

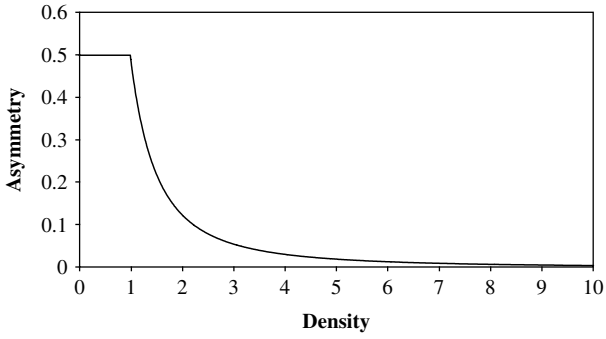

Fig. 1. Prior density of the asymmetry parameter (γ).

(iv) *Markov Chain Monte Carlo algorithm*

A Gibbs sampling procedure (Gelfand & Smith, 1990) was performed to obtain samples from marginal

posterior distributions of the parameters. The Gibbs sampling procedure involves an updating sampling scheme of the fully conditional distributions of all unknown parameters involved in the model. Here, the conditional distributions of each level of the systematic effects was univariate normal, and the conditional distribution of the residual variance (σ_e^2) was a scale inverted chi-square distribution. Sampling of elements of the random incidence matrix (**M**) was performed by drawing each element from a discrete probability distribution in the following way:

$$p(M_{ij}=k|\beta,\mathbf{q},\sigma_e^2,\mathbf{y},Mks) = \frac{f(\mathbf{y}|\beta,\mathbf{q},\sigma_e^2,M_{ij}=k)f(M_{ij}=k|Mks)}{\sum_{k=-1.0,1} f(\mathbf{y}|\beta,\mathbf{q},\sigma_e^2,M_{ij}=k)f(M_{ij}=k|Mks)}$$

where $f(M_{ij}=k|\beta,\mathbf{q},\sigma_e^2,\mathbf{y})$ is the conditional distribution of the elements of $\mathbf{M}=\{-1,0,1\}$ given the data, systematic and marker effects and the residual variance. In contrast, sampling from the conditional distribution for each level of the marker effects (q_i), the variance of the marker effects (σ_q^2), the asymmetry parameter (γ) and the degrees of freedom of the *t*-distribution (ν) was performed using a Metropolis–Hastings algorithm (Hastings, 1970) with a uniform proposal distribution with a range of 10 units around the value of the previous iteration.

Based on output from the Gibbs sampler, the percentage of variance explained by the marker effects (h_q^2) was calculated as

$$h_q^2 = \frac{\sum_{i=1}^{n_m} 0{\cdot}5q_i^2}{\sum_{i=1}^{n_m} 0{\cdot}5q_i^2 + \sigma_e^2}$$

since the additive variance explained by the *i*th marker is $0{\cdot}5q_i^2$.

We also calculated an 'empirical marginal prior' distribution of the *i*th marker effect evaluated by averaging the conditional prior distributions in (1) over the posterior samples for σ_q^2 and γ at each iteration by using the Rao–Blackwell argument:

$$f(q_i) = \sum_{j}^{niter} f(q_i|\sigma_{q_j}^2,\gamma_j)/niter.$$

Above, *niter* is the total number of iterations after convergence, and σ_{qj}^2 and γ_j are the variance of the additive genetic marker effects and the asymmetry parameter, respectively, sampled in the *j*th iteration. The objective of this empirical marginal prior distribution is to represent the prior influence on the marker effects.

The Gibbs sampling analysis was performed by running 110 000 iterations, with the first 10 000 discarded as burn-in. Convergence was checked using the CODA software (Best *et al.*, 1995). All 100 000 samples were used for extracting posterior features.

4. Results

(i) *Live weight*

The distribution of posterior means of additive marker effects with the Gaussian and the asymmetric *t* prior distributions for live weight are presented in Fig. 2 (top panel). The most extreme effect was the −1·01 kg associated with marker SW24 in SSC17. The mean of the posterior distribution of the asymmetry parameter (γ) was 1·00, and the posterior standard deviation was 0·08 (Fig. 3, top panel). These results imply that the probability mass below 0 in the empirical marginal prior distribution of the marker effects was 0·50, as can be observed in Fig. 4 (top panel). In addition, the posterior mode of the degrees of freedom of the *t*-distribution was 16, and the probability that the degrees of freedom were larger than 30 was 52·59 % (Fig. 5, top panel). Moreover, the posterior mean estimate of the variance pertaining to the effects associated with the markers was 0·411 (0·144), and the posterior mean estimate of residual variance was 45·64 (4·63). Finally, the posterior mean of the proportion of variance explained by the markers in the F_2 population was 0·441 (0·094).

(ii) *Carcass length*

Results for posterior mean estimates of the additive marker effects with the Gaussian and asymmetric *t* prior distributions on carcass length are presented in Fig. 2 (centre panel). The effects with the highest absolute values were associated with markers S0001 (−0·28 cm), SW2571 (−0·31 cm) and SW24 (−0·40 cm), located in SSC4, SSC9 and SSC17, respectively. The posterior distribution of γ for carcass length is presented in Fig. 3 (centre panel), and had a mean of 1·27 and a standard deviation of 0·15. The posterior probability of γ being below 1·00 was only 1·02 %. These results imply that the probability mass below 0 in the empirical marginal prior distribution of the marker effects was 0·61 (Fig. 4, centre panel). In addition, the posterior distribution of the degrees of freedom of the *t*-distribution had a mode at 13, and the posterior probability of degrees of freedom being greater than 30 was 45·26 % (Fig. 5, centre panel). The posterior mean estimate of the variance of the effects associated with the markers was 0·026 (0·014) and the posterior mean estimate of the residual variance was 4·59 (0·48). The posterior mean of the proportion of variance explained by the markers in the F_2 population was 0·327 (0·121).

(iii) *Backfat depth*

Posterior mean estimates of the additive marker effects with Gaussian and asymmetric *t* prior distributions are presented in Fig. 2 (bottom panel). The

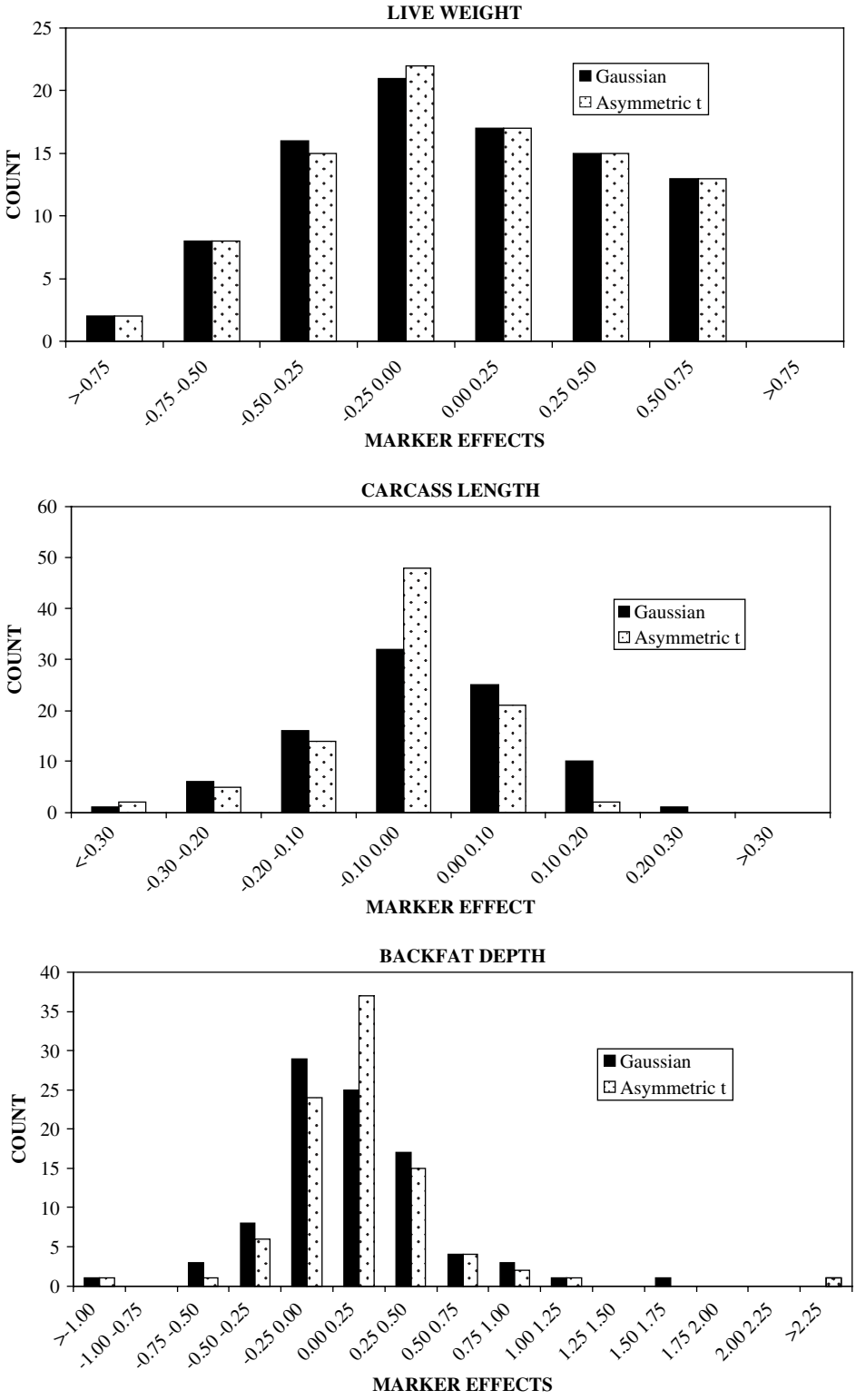

Fig. 2. Posterior mean estimates with the Gaussian and asymmetric *t* models for marker effects for live weight, carcass length and backfat depth.

most sizeable additive effects were associated with markers S0001 (+0·99 mm) in SSC4, SW316 (+1·09 mm) and S0228 (+2·30 mm) in SSC6, and S0101 (–1·05 mm) in SSC7. The posterior mean and standard deviation of γ were 0·87 and 0·08, respectively (Fig. 3, bottom panel), with a posterior probability of γ greater than 1 equal to 4·7%. As a consequence, the probability mass below 0 in the

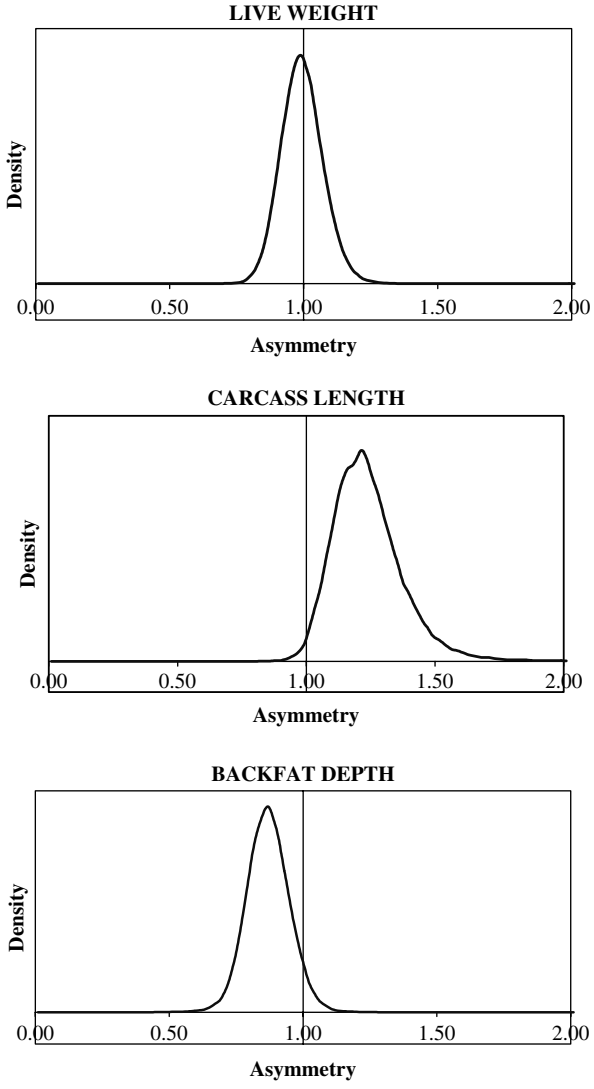

Fig. 3. Posterior density of the asymmetry parameter (γ) for live weight, carcass length and backfat depth.

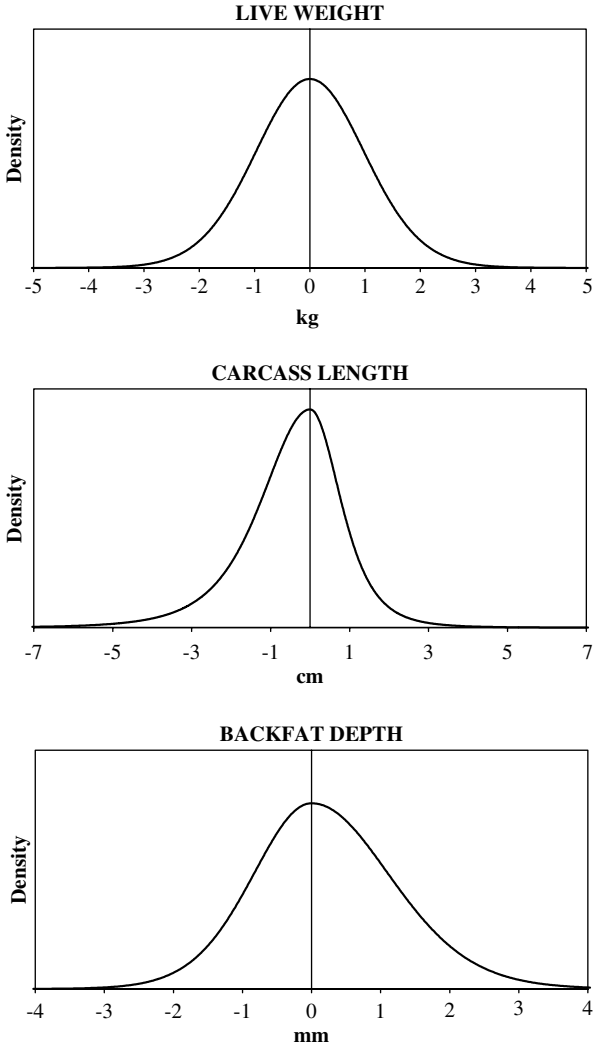

Fig. 4. Empirical marginal prior density of marker effects for live weight, carcass length and backfat depth.

empirical marginal prior distribution of the marker effects was 0·43 (Fig. 4, bottom panel). In Fig. 5 (bottom panel), we present the posterior distribution of the degrees of freedom of the *t*-distribution, with a posterior mode of 2, and a probability of the degrees of freedom being greater than 30 equal to only 17·98 %. The posterior mean estimate of the variance of the effects associated with the markers was 0·195 (0·120), and the posterior mean estimate of the residual variance was 22·94 (2·31). The posterior mean of the proportion of variance explained by the markers in the F_2 population was 0·397 (0·170).

5. Discussion

The *t*-distribution has been widely used to model deviations from Gaussian assumptions (Lange *et al.* 1989), even in the field of quantitative genetics (Stranden & Gianola, 1999; Rosa *et al.*, 2004). However, these *t*-distributions have been used mostly to model residuals, due to the suspect unknown non-random preferential treatment associated with some observations (Stranden & Gianola, 1999). In our case, we used it to model the prior distribution of effects associated with molecular markers. The *t*-distribution can account better than the Gaussian distribution for heavy tails, and its use is justified on the basis of evidence suggesting that some markers or segments of the genome are strongly associated with major genes or QTL. In the present study, we replaced the prior Gaussian distribution with a robust *t*-distribution in the procedure designed by Gianola *et al.* (2003).

Furthermore, the symmetric distribution of Gianola *et al.* (2003) was replaced by an asymmetric distribution (Fernandez & Steel, 1998), which allows for differences in the density of positive and negative effects. Most experiments for detecting QTL are based on F_2 designs involving crosses between divergent lines. With this scheme, most of the QTL effects are

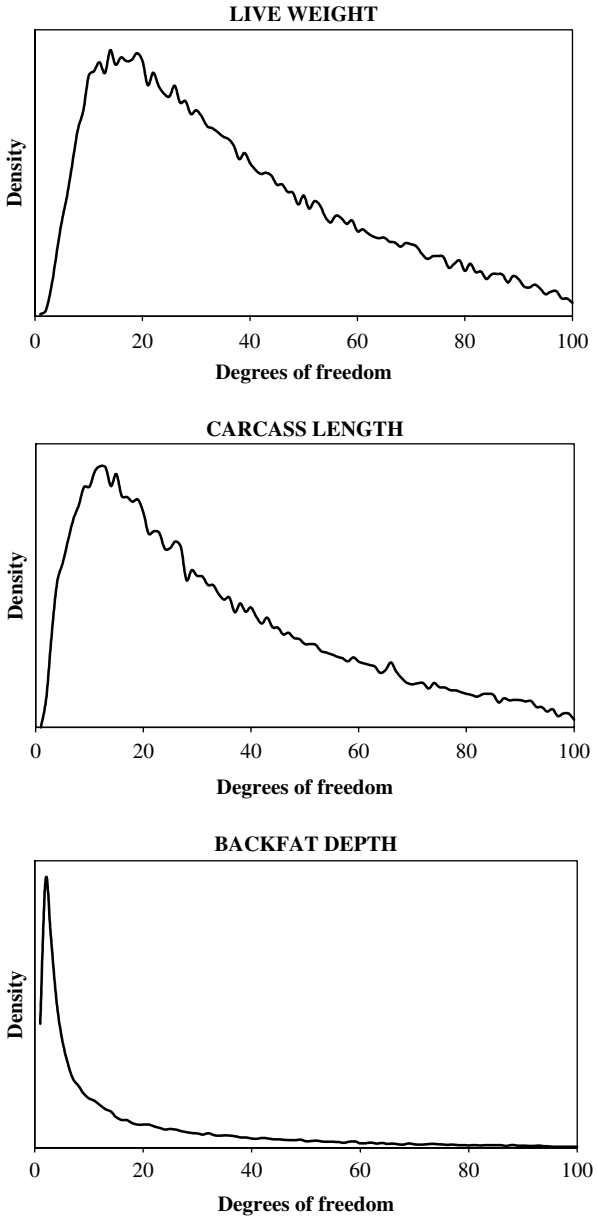

Fig. 5. Posterior density of the degrees of freedom (v) of the t-distribution for live weight, carcass length and backfat depth.

expected to favour the most productive line, whereas only a few are expected to be cryptic, or to favour the lowest production line. Therefore, the asymmetric distributions described by Fernandez & Steel (1998) can be very useful for modelling asymmetry of effects.

It must be noted that the symmetric Gaussian prior defined by Gianola *et al*. (2003) is a particular case of the proposed model. The asymmetric robust distribution converges to the Gaussian distribution when the asymmetry parameter (γ) is equal to 1 and when the degrees of freedom of the t-distribution are large enough.

One of the most controversial points in Bayesian analysis is the specification of prior distributions. Here, the prior distribution of γ (Fig. 1) was assumed flat between 0 and 1 with a probability of 0·5, and with a density equal to $\frac{0.5}{\gamma^2}$ when γ is greater than 1. This distribution ensured that 50% of prior probability was below $\gamma = 1$ and 50% over $\gamma = 1$ and, in addition, that the probability of a subspace between 0 and 1 was equal to the probability between the reciprocal of the boundaries of the subspace. For example, the prior probability of γ being between 0·3 and 0·5 was 0·1, which is the same as the probability of it being between $\frac{1}{0.3}$ and $\frac{1}{0.5}$. In addition, the prior distribution of the degrees of freedom was assumed to be uniform between values of 1 and 100. The upper limit of 100 was established by assuming that differences between t-distributions with degrees of freedom over 100 are negligible, so the resulting formula would be equivalent to that of a univariate Gaussian distribution.

In the case analysed, we presented results for three traits, leading to different pictures with respect to asymmetry and degrees of freedom of the t-distribution. The posterior distribution of γ was centred on 1 for live weight, whereas the probability of γ being over 1 was higher for carcass length and lower for backfat depth (Fig. 3). These results support the idea that the empirical marginal prior distribution of the marker effects should be asymmetric for backfat depth and carcass length (Fig. 4). As a consequence, such an empirical marginal prior distribution implies an increase in the proportion of negative effects on carcass length relative to a Gaussian prior distribution (Fig. 2, centre panel). This would favour the Landrace line, due to the differences between the two breeds reported in the literature (Serra *et al*., 1998). Similarly, an increase in the proportion of positive effects (favouring the Iberian line) was observed for backfat depth (Fig. 2, bottom panel). Again, these results are consistent with differences between the parental breeds reported in the literature. For instance, Serra *et al*. (1998) reported that mean backfat was 48 mm for the Iberian and 20 mm for the Landrace pigs, respectively.

With respect to the degrees of freedom of the t-distribution, posterior distributions for live weight and carcass length had modes, means and medians close to 20, while the probability of their respective degrees of freedom being over 30 was around 50%. These results imply that the distribution of marker effects does not differ substantially from a Gaussian distribution, and that including a robustness correction is not strictly needed. On the other hand, for backfat depth, the posterior distribution of the degrees of freedom had a mode of 2, and the posterior probability of being over 30 was only 17·98%. Here, the robustness 'correction' was useful, and these results agree with those presented by Varona *et al*. (2002), showing that some QTL with major effects

could explain differences between the two lines. These QTL effects associated with markers S0001 (+0·99 mm) in SSC4, SW316 (+1·09 mm) and S0228 (+2·30 mm) in SSC6, and S0101 (−1·05 mm) in SSC7.

The case studied provided a simple description of differences between two line origins in an F_2 experiment. Marker effects for live weight were appropriately described with a Gaussian prior distribution, whereas marker effects for carcass length were clearly asymmetric, and for backfat depth a skewed *t*-distribution seemed more appropriate.

The 92 marker effects were assumed to be drawn from a single distribution, but extensions to this model can be implemented with ease. Marker effects could be split into between- and within-chromosome deviations, as suggested by Gianola *et al.* (2003). Moreover, the evidence of co-expression in adjacent regions of the genome (Caron *et al.*, 2001) suggests a covariance structure between the various marker effects. In this sense, multivariate skew distributions (Gupta, 2003; Gupta *et al.*, 2004) could be used to accommodate relationships between markers in the prior distribution, and even to estimate the relationship parameter(s) via a Metropolis–Hastings algorithm (Hastings, 1970). In addition, background polygenic effects could also be modelled by assuming either a multivariate normal distribution with the numerator relationship matrix (Quaas, 1976), or with multivariate *t*-distributions as proposed by Stranden & Gianola (1999), and the segregation variance could be modelled as suggested by Birchmeier *et al.* (2002).

The procedure proposed by Xu (2003) differs from that proposed by Gianola *et al.* (2003) in the definition of the prior distributions. The former author proposed a prior distribution unique to each of the marker effects, whereas Gianola *et al.* (2003) argued that marker effects are a sample from a more general prior distribution. A skewed *t*-distribution can also be fitted in the procedure of Xu (2003), by assuming that the asymmetry parameter (γ) and the degrees of freedom (t) are common to all prior distributions of the marker effects.

In conclusion, as illustrated in this study, the skewed *t* prior distribution proposed offers flexibility and robustness concerning the distribution of genetic effects associated with molecular markers. The use of shrinkage as suggested by Whittaker (2001) and Lange & Whittaker (2001), and generalized into a Bayesian setting by Gianola *et al.* (2003), avoids problems due to overparameterization, and colinearity causing unstable least-squares estimates. The skewed *t*-distribution retains the desirable properties described in Gianola *et al.* (2003) but also makes it possible to describe differences between lines in an F_2 cross more adequately. This procedure can also be used to make a global description of differences between parental lines due to additive, dominance and epistatic effects.

Some other possible extensions of the procedure can be considered. First, the number of degrees of freedom can vary between positive and negative effects, accounting for possible differences in heavy tails favouring each of the founder lines. Second, the residual distribution considered here was Gaussian, but it can be replaced by a skewed *t*-distribution, following Von Rohr & Hoeschele (2002). In addition, further research must be done to compare the proposed model with other alternatives for modelling asymmetry of effects, such us the use of mixtures of distributions (Gianola *et al.*, 2006).

Acknowledgements

We would like to thank all the participants in the Spanish CICYT grants AGF96-2510-C05, AGF99-0284-C02, SC00-057 and CPE03-010 for making data available.

References

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. & Lundstrom, K. (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**, 1771–1774.

Best, N., Kathryn, M. K. & Vines, K. (1995). CODA: convergence diagnosis and output analysis software for Gibbs sampling output. Version 3.0. Cambridge: Medical Research Council Biostatistics Unit.

Birchmeier, A. N., Cantet, R. J. C., Fernando, R. L., Morris, C. A., Holgado, F., Jara, A. & Cristal, M. S. (2002). Estimation of segregation variance for birth weight in beef cattle. *Livestock Production Science* **76**, 27–35.

Caron, H., Van Schaik, B., Van der Mee, M., Baas, F., Riggins, G., Van Sluis, P., Hermus, M. C., Van Asperen, R., Boon, K., Voute, P. A., Heisterkamp, S., Van Kampen, A. & Versteeg, R. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289–1292.

DeVicente, M. C. & Tanksley, S. D. (1993). QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* **134**, 585–596.

Fernandez, C. & Steel, M. F. J. (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* **93**, 359–371.

Gelfand, A. & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gianola, D., Perez-Enciso, M. & Toro, M. A. (2003). On marker-assisted prediction of genetic values: beyond the ridge. *Genetics* **163**, 347–365.

Gianola, D., Heringstad, B. & Odegaard, J. (2006). On the quantitative genetics of mixture characters. *Genetics* **173**, 2247–2255.

Gupta, A. K. (2003). Multivariate skew *t*-distribution. *Statistics* **37**, 359–363.

Gupta, A. K., Gonzalez-Farias, H. & Domínguez-Molina, J. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis* **89**, 181–190.

Haley, C. S., Knott, S. A. & Elsen, J. M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207.

Hastings, W. K. (1970). Monte-Carlo methods using Markov chains and their applications. *Biometrika* **57**, 97.

Iannucelli, E., Wolosyn, N., Arhainx, J., Gellin, J. & Milan, D. (1996). Gemma: a database to manage and automate microsatellite genotyping. In *Proceedings of the International Society of Animal Genetics Conference*, Tours, France, p. 88.

Lander, E. S. & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Lange, C. & Whittaker, J. C. (2001). On prediction of genetic values in marker-assisted selection. *Genetics* **124**, 743–756.

Lange, K. L., Little, R. J. A. & Taylor, J. M. H. (1989). Robust statistical modelling using the *t*-distribution. *Journal of the American Statistical Association* **84**, 881–896.

Mekkawy, W. (2005). QTL bayesian analysis. Doctoral thesis, Universidad Politecnica de Valencia.

Pe, M. E., Gianfranceschi, L., Taramino, G., Tarchini, R., Angelini, P., Dani, M. & Binelli, G. (1993). Mapping quantitative trait loci (QTLs) for resistance to *Gibberella zeae* infection in maize. *Molecular and General Genetics* **241**, 11–16.

Quaas, R. L. (1976). Computing diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* **32**, 949–953.

Ron, M., Lewin, H., Da, Y., Band, M., Yanai, A., Blank, Y., Feldmesser, E. & Weller, J. I. (1995). Prediction of informativeness for microsatellite markers among progeny of sires used for detection of economic trait loci in dairy cattle. *Animal Genetics* **26**, 439–441.

Rosa, G., Gianola, D. & Padovani, C. R. (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *Journal of Applied Statistics* **31**, 855–873.

Serra, X., Gil, M., Pérez-Enciso, M., Oliver, M. A., Vazquez, J. M., Gispert, M., Diaz, I., Moreno, F., Latorre, J. L. & Noguera, J. L. (1998). A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. *Livestock Production Science* **56**, 215–223.

Stranden, I. & Gianola, D. (1999). Mixed effects linear models with *t*-distributions for quantitative genetic analysis: a Bayesian approach. *Genetics Selection Evolution* **31**, 25–42.

Toro, M. A., Rodrigáñez, J., Silió, L. & Rodríguez, C. (2000). Genealogical analysis of a closed herd of black hairless Iberian pigs. *Conservation Biology* **14**, 1843–1851.

Varona, L., Ovilo, C., Clop, A., Noguera, J. L., Pérez-Enciso, M., Coll, A., Folch, J. M., Barragán, C., Toro, M. A., Babot, D. & Sanchez, A. (2002). QTL mapping for growth and carcass traits in an Iberian by Landrace pig intercross: additive, dominant and epistatic effects. *Genetical Research* **80**, 145–154.

Von Rohr, P. & Hoeschele, I. (2002). Bayesian QTL mapping using skewed Student *t*-distributions. *Genetics Selection Evolution* **34**, 1–21.

Whittaker, J. C. (2001). Marker-assisted selection and introgression. In *Handbook of Statistical Genetics* (ed. D. J. Balding, M. Bishop & C. Cannings), pp. 673–693. Chichester: Wiley.

Xu, S. (2003). Estimating polygenic effects using markers of the entire genome. *Genetics* **163**, 789–801.