

Contents lists available at [ScienceDirect](http://ScienceDirect)

## Livestock Science

journal homepage: [www.elsevier.com/locate/livsci](http://www.elsevier.com/locate/livsci)

## Three-step Bayesian factor analysis applied to QTL detection in crosses between outbred pig populations

Fabyano Fonseca Silva <sup>a,b,\*</sup>, Guilherme J.M. Rosa <sup>b</sup>, Simone E.F. Guimarães <sup>c</sup>, Paulo Sávio Lopes <sup>c</sup>, Gustavo de los Campos <sup>d</sup>

<sup>a</sup> Department of Statistics, Universidade Federal de Viçosa, Av. P.H. Hofls, 36570-000, Viçosa, Brazil

<sup>b</sup> Department of Animal Science, University of Wisconsin, 460 Animal Science Building, Madison, USA

<sup>c</sup> Department of Animal Science, Universidade Federal de Viçosa, Av. P.H. Hofls, Viçosa, Brazil

<sup>d</sup> Biostatistics Department, University of Alabama, 1665 University Boulevard, Birmingham, USA

### ARTICLE INFO

#### Article history:

Received 10 March 2011

Received in revised form 24 June 2011

Accepted 23 July 2011

#### Keywords:

QTL regression method  
Data dimension reduction  
Pig microsatellite markers

### ABSTRACT

Marker assisted selection (MAS) can be used to improve the efficiency of genetic selection of traits for which phenotypic measurements are expensive or cannot be obtained on selection candidates, such as carcass traits. Marker information required for MAS may be acquired through the identification of QTLs. Generally, univariate models are used for QTL detection, although multiple-trait models (MTM) may enhance QTL detection and breeding value estimation. In MTM, however, the number of parameters can be large and, if traits are highly correlated, such as carcass traits, estimates of (co) variance matrices may be close to singular. Because of this, dimension reduction techniques such as Factor Analysis (FA) may be useful. The aim of our project is to evaluate the use of FA for structuring (co)variance matrices in the context of Bayesian models for QTL detection in crosses between outbred populations. In our method, QTL effects are postulated at the level of common factors (CF) rather than the original traits, using a three-step approach. In a first step, a MTM is fitted to arrive at estimates of systematic effects and prediction of breeding values (procedure A) and only systematic effect (procedure B). These estimates/predictions are then used to generate an adjusted phenotype that is further analyzed with a Bayesian FA model. This step yields estimates of factor scores for each animal and CF. In the last step, the scores relative to each CF are analyzed independently using probabilities for the line of origin combination. To illustrate the methodology, data on 416 F2 pigs (Brazilian Piau X commercial) with ten traits (5 fat-related, 2 loin measurements, and 3 carcass classification systems) were analyzed. For each of the three resulting CFs, an independent QTL scan was performed on chromosome 7 considering three models: I) null (i.e., absence of QTL); II) additive effect QTL, and III) additive and dominance effect QTL. The posterior probability (PP) of each model was calculated from Bayes factor for each considered procedures (A and B). A Three-step Bayesian factor analysis allowed us to calculate the probability of QTLs that simultaneously affect a group of carcass traits for each position of SSC 7. The removal of systematic effects in the first step of the evaluation (procedure B) allowed that the factor analysis, which was performed in the second step, identify three distinct factors that explained 85% of the total traits variation. For the common factor that represented fat-related traits (bacon depth, midline lower backfat thickness, higher backfat thickness on the shoulder; midline backfat thickness after the last rib; midline backfat thickness on the last lumbar vertebrae) the third step of the analysis showed that the highest probability of an additive QTL effect at the 65 cM position was 86%.

© 2011 Elsevier B.V. Open access under the [Elsevier OA license](http://Elsevier OA license).

\* Corresponding author at: Departamento de Estatística, UFV, Av. P.H.Hofls, 36570-000, Viçosa, Minas Gerais, Brasil. Tel.: +55 031 38991790; fax: +55 031 38993880.  
E-mail addresses: [fabyanofonseca@ufv.br](mailto:fabyanofonseca@ufv.br) (F.F. Silva), [grosa@wisc.edu](mailto:grosa@wisc.edu) (G.J.M. Rosa), [sfacioniguima@gmail.com](mailto:sfacioniguima@gmail.com) (S.E.F. Guimarães), [pslopos@ufv.br](mailto:pslopos@ufv.br) (P.S. Lopes), [gcampos@ms.soph.uab.edu](mailto:gcampos@ms.soph.uab.edu) (G. de los Campos).

## 1. Introduction

Currently, to meet the growing and increasingly demanding consumer market, characterized by higher meat yield and lower external fat deposition, carcass characteristics are very important for the development of pig farming. Thus, genetic improvements that are carried over successive generations are an indispensable tool for obtaining animals with such desirable characteristics.

A considerable reduction of the number of generations that are needed to produce animals with the required carcass standard can be obtained through the association of phenotypic, given by carcass quality measures, and genotypic information, being the latter characterized by genetic markers that can be linked to loci that affect quantitative traits (QTL's). Thus, several studies have been developed to identify QTLs that are related to carcass traits in pigs (Demeure et al., 2005; Milan et al., 2002; Wang et al., 1998; Yue et al., 2003).

Recent genome scanning revealed that there are genes located within quantitative trait locus (QTL) regions for fat related traits, including regional backfat thickness and intramuscular fat content (Szydlowski et al., 2011), but other traits like the quality of the cuts and carcass classification systems also can be important. In view of this larger number of traits, multivariate models to detect QTLs can be useful, however, the direct application of these models can present problems, such confused interpretations of the results and singularity of the (co)variance matrices, which are due to the strong correlations between some traits.

A practical method for solving these problems is to transform the original variables so that they become uncorrelated, i.e., the use of techniques based in dimension data reduction, such as principal component (PCA) and factor analysis (FA). Under these approaches, is possible to detect QTLs associated whit latent variables, components or factors, which represent several traits simultaneously.

Stearns et al. (2005) used the PCA technique to analyze QTLs for growth, carcass, and meat quality traits in pigs from the F2 generation of a Berkshire × Duroc. The authors reported the advantages of working with a small number of parameters and of using univariate methods to detect QTLs for the orthogonal components. Although the PCA technique is relevant, it has some disadvantages; it is restricted to the condition of orthogonality between, it is susceptible to changes in the scales and lacks an adequate criterion to determine when a sufficient proportion of the total variation is explained by the retained components (Johnson and Wichern, 2007).

FA is a more appropriate technique in which each variable response is represented by a linear function of a small number of non-observable common factors and a simple latent specific variable. The coefficients of the common factors are not restricted to the condition of orthogonality, which confers generality but requires normality of the data and determination of the number of factors a priori (Johnson and Wichern, 2007). These last two conditions of FA can be easily dismantled under the Bayesian approach, which has been highly effective in the area of animal breeding during the last decade. Bayesian FA was employed successfully by de los Campos and Gianola (2007) in assessments of the multi-characteristic mixed model based on pedigree information (Henderson, 1984). However, to date, there have been no reports concerning the application of this Bayesian technique in multivariate analysis of QTLs.

## 2. Objectives

The aim of the present paper is to evaluate the use of FA for structuring (co)variance matrices in the context of Bayesian models for QTL detection in crosses between outbred F2 population in pigs Brazilian Piau X commercial. In the proposed method, QTL effects are postulated at the level of common factors (CF) rather than the original traits, using a three-step Bayesian approach.

## 3. Materials and methods

### 3.1. Population assessment, phenotypic data and genotypic information

The F2 pig population was generated by crossing two native Brazilian Piau boars with 18 commercial sows (Landrace × Large White × Pietran) selected for growth rate and backfat thickness. The F1 generation consisted of 106 sows and 134 boars (Band et al., 2005). Twelve boars from different litters were randomly selected from the 134 F1 boars and mated by natural breeding with 54 F1 sows to produce the F2 generation. The F2 generation consisted of approximately 840 offspring divided into five batches according to the season in which they were born.

After slaughter, around 65 kg of living weight ( $64.71 \pm 0.24$ ), the following carcass traits were evaluated in the animals of F2 generation: bacon depth (BCD), midline lower backfat thickness (L), higher backfat thickness on the shoulder (SBT); midline backfat thickness after the last rib (LR); midline backfat thickness on the last lumbar vertebrae (LL), loin eye area (LEA), loin depth (LD), Brazilian (MBCC) and American (MLC) carcass classification method, and carcass yield (CY).

Details of the DNA extraction procedures used have been described by Faria et al. (2006). Six primer pairs for microsatellite markers distributed on SSC7 (S0025, S0064, S0102, SW252, SW632, and S0212) were used. Amplifications were done in an MJ Research PTC 100-96® thermocycler, according to standard laboratory procedures. The amplified fragments were scored automatically by GenScan software installed in an ABI PRISM 310 sequencer (Applied Biosystems). Annotation and genotype checking were done manually by two independent and previously trained technicians. The CRIMAP software (Lander and Green, 1987) was used to construct linkage maps of the related markers, which were distributed, respectively, at positions 0, 31, 65, 96, 108 and 136 cM.

### 3.3. Bayesian factor analysis

Briefly, FA is a statistical technique that is used to reduce the size of a multivariate data set. This is accomplished by the identification of common factors that originate from (co) variance measurements among the considered variables. According to Meyer (2009),  $\mathbf{y}$  is a vector of  $q$  random variables with a (co)variance matrix that is given by  $\Sigma$ , and therefore, the following model can be considered:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{c} + \mathbf{s}, \quad (1)$$

where  $\boldsymbol{\mu}$  is a vector of the average;  $\boldsymbol{\Gamma}$  is a matrix called "factorial loadings" with dimensions of  $q \times m$  such that  $m \leq q$ ;

$\mathbf{c}$  is a vector of “common factors” with dimensions of  $m \times 1$ ; and  $\mathbf{s}$  is a vector of the residuals, which are also called “specific factors”, with dimensions of  $q \times 1$ . In the simplest case involving FA, the columns ( $\gamma_i$ ) of  $\Gamma$ , which are given by the square root of the eigenvectors of  $\Sigma$ , which are considered orthogonal. That is,  $\gamma_i' \gamma_j = 0$  for  $i \neq j$ , such that the elements of  $\mathbf{c}$  are not correlated. Furthermore, the elements of  $\mathbf{c}$  are assumed to present a unit variance,  $V(\mathbf{c}) = \mathbf{I}$ , and independent of  $\mathbf{s}$ . For this last vector, the variance of the  $i$ th element is given by  $\Psi_i$ .

According to the presented theory,  $\Sigma$  can be decomposed as  $\Sigma = \Gamma' \Gamma + \Psi$ , where  $\Psi = \text{diag}(\Psi_i)$ . In this representation, all of the (co)variances between the variables in  $\mathbf{y}$  appear to be explained by some common factors ( $\mathbf{c}$ ). The specific factors ( $\mathbf{s}$ ) describe the additional variance that is restricted to each of these variables. In general, for  $m$  common factors, the  $q(q + 1)/2$  elements of  $\Sigma$  can be represented by  $q + mq - m(m - 1)/2$  parameters obtained from the FA. Thus, for small values of  $m$ , the FA provides a parsimonious modeling of  $\Sigma$ .

The parameters of the model presented in (1) can be estimated using different methods, and the Bayesian inference was utilized herein. For this purpose, we developed two different procedures in which three distinct steps were considered. For the procedure A, in the first step a Bayesian analysis of a multivariate mixed model was performed with the goal of removing polygenic and environmental effects, such that the scores resulting of the Bayesian factors analysis in the second step depend solely on the QTL effects. Thus, in the third step, the estimated factor scores from the second step were considered as dependent variables in a Bayesian linear regression to estimate the QTL effects. For the procedure B, in the first step the multivariate model contemplated only environmental effects, thus in the second step the resulting scores depended on the QTL and polygenic effects, which were embedded in the last step model. In summary, the procedures A and B comparisons are related to hypothesis that the estimation of QTL effects simultaneously with polygenic effects (procedure B) can produce most significant results, because a relevant percentage of variance explained by the QTL may be removed by the polygenic effects in the first step of the procedure A.

In the first step of procedure A, the following multivariate mixed model was adopted:

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

where  $y = (y_1, y_2, \dots, y_p)'$  is the vector for the observation of  $p$  variables,  $\beta$  and  $\mathbf{u}$  are the vectors for systematic effects (sex and batch) and polygenic effects, respectively, and  $\mathbf{e}$  is a vector for random errors.

Using the Bayesian approach, the following distribution of the sample data was assumed:  $y|\beta, u, G, R \sim N(X\beta + Zu, I \otimes R)$ ; and non-informative priors (flat) was attributed for  $\beta$  and for the (co)variance components ( $\mathbf{G}$  and  $\mathbf{R}$ ). For  $\mathbf{u}$  the adopted distribution was:  $u|A, G \sim N(0, A \otimes G)$ . After determining estimates for the parameters of the model in question, we obtained  $w = \hat{e} = y - (X\hat{\beta} + Z\hat{u})$ . The free software GIBBSF90 (Misztal et al., 2002) was used for this analysis, and a total of 150,000 iterations, with burn-in of 50,000 and thin of 2 iterations were

considered. For the procedure B, the model (2) did not consider the polygenic effect term.

In the next step, the adjusted traits of each animal ( $w_i$ ) were reduced to common factors ( $c_i$ ) in accordance with the following factor model:

$$w_i = \Gamma c_i + s_i, \quad (i = 1, 2, \dots, n); \tag{3}$$

where  $\Gamma = (\lambda_{jk})$  is a matrix of the factorial loadings  $p \times q$  ( $j = 1, 2, \dots, p$  and  $k = 1, 2, \dots, q$ );  $\mathbf{c}_i = (c_{i1}, \dots, c_{iq})'$  is the vector of dimension  $q \times 1$  for common factors determined for animal  $i$ ;  $\mathbf{s}_i = (s_{i1}, \dots, s_{ip})'$  is the vector of dimension  $p \times 1$  for specific factors determined for animal  $i$ .

Given the model in question, the following distribution for the sample data was assumed:  $\begin{pmatrix} c_i \\ s_i \end{pmatrix} \stackrel{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_q & 0 \\ 0 & \Psi \end{pmatrix}\right)$ ,

where  $\Psi = \text{diag}(\psi_j)$ . Therefore,  $w_i \stackrel{iid}{\sim} N(0, \Sigma)$ , with matrix  $\Sigma$  given by  $\Sigma = \Gamma \Gamma' + \Psi$ . The prior distributions for the parameters of the factorial model were as follows:  $\lambda_{jk} \sim N(\mu_{\lambda_{jk}}, \sigma_{\lambda_{jk}}^2)$ ,  $c_i \sim N(0, I_q)$ ,  $e \sim IG(a, b)$ . To implement this Bayesian analysis the *MCMCfactanal* function available in the *MCMCpack* of the free R software (R Development Core Team, 2010) was used considering the same number of iteration of the previous step. The assumed values for the hyperparameters (prior distribution parameters) were given by the *MCMCfactanal* function default.

In the third and final step of the analysis, the estimated factor score for each animal ( $\hat{c}_i$ ), which was given by the posterior mean of  $c_i$ , was used as the phenotypic characteristics in a Bayesian linear regression model as proposed by Haley et al. (1994).

For the procedure A, in this model the independent variables are the coefficients  $a$  (additive effect) and  $d$  (dominance effect), which are related to the probabilities for the line of origin combination. These coefficients are expressed as follows:  $a = P(QQ) - P(qq)$  and  $d = P(Qq)$ , where  $P(QQ)$ ,  $P(qq)$ , and  $P(Qq)$  are the probabilities of the alleles being commercial homozygous, Piau homozygous, and heterozygous, respectively. The regression model in question is given by:

$$\hat{c} = Q\alpha + \varepsilon;$$

where  $\hat{c} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)'$  is the vector for the estimated factor scores of  $n$  individuals,  $Q$  is the incidence matrix of the effects represented by  $\alpha$ , and  $\varepsilon$  is the vector of random errors. Under this approach, three different models were considered, and each model had its own particular incidence matrix.

Model I is the null model (presence of the intercept only), model II includes additive QTL effects, and model III includes additive and dominance QTL effects. The incidence matrices for each of these models were, respectively:

$$Q_I = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}; \quad Q_{II} = \begin{bmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix}; \quad Q_{III} = \begin{bmatrix} 1 & a_1 & d_1 \\ \vdots & \vdots & \vdots \\ 1 & a_n & d_n \end{bmatrix},$$

where the values for coefficients  $a_i$  and  $d_i$  of each animal  $i$  were obtained from *Qxpack* software (Pérez-Enciso and Misztal, 2004). To perform the Bayesian inference, the following distributions

were adopted:  $\hat{c}|\alpha, \sigma^2 \sim N(Q\alpha, \sigma^2 I_n)$ ,  $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2 I_3)$  and  $\sigma^2 \sim IG(c, d)$ . For this purpose, we used the *MCMCreg* function available in the *MCMCpack* of the R software (R Development Core Team, 2010). The iterations number, bur-in and thin were the same of the first and second step. In order to compare the models I, II, and III and thus infer about the presence of QTLs in each chromosome 7 position (see Fig. 1), we calculated the model posterior probability using the *PostProbMod* function available in the *MCMCpack* of the R software (R Development Core Team, 2010). This probability is provided by the following expression:

$$p(M_t|\hat{c}) = \frac{p(\hat{c}|M_t)p(M_t)}{\sum_{t=I,III,III} p(\hat{c}|M_t)p(M_t)}, t = I, III, III;$$

where  $p(\hat{c}|M_t)$  is the marginal likelihood calculated by Newton and Raftery (1994) method, and  $p(M_t)$  is the model prior probability, assumed as the same for all models.

For the procedure B, in the last step the following model was considered:

$$\hat{c} = Zu + Q\alpha + \varepsilon;$$

where  $u$  is the polygenic effect with incidence matrix given by  $Z$ . The following distributions were used in the Bayesian analysis of this model:  $\hat{c}|u, \sigma_u^2, \alpha, \sigma^2 \sim N(Zu + Q\alpha, \sigma^2 I_n)$ ,  $u|\sigma_u^2 \sim N(0, A\sigma_u^2)$ ,  $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2 I_3)$ ,  $\sigma^2 \sim IG(c, d)$  and  $\sigma_u^2 \sim IG(e, f)$ . In order to fit the model in question, the package *MCMCgmm* (with the option *pedigree*) of the R software was used, being the iterations number, bur-in and thin the same as those previously presented. Under this procedure B, in order to compare the models I, II, and III, that now contain the polygenic effect, the model posterior probabilities were calculated from DIC (Spiegelhalter et al., 2002) using the approximation presented by Wilberg and Bence (2008), which is given by:

$$p(M_t|\hat{c}) = \frac{\exp(-\Delta_t/2)}{\sum_{t=I,III,III} \exp(-\Delta_t/2)}, t = I, III, III;$$

where:  $p(M_t|\hat{c})$  is the posterior probability of model  $t$ ,  $\Delta_t$  is the DIC difference between model  $t$  and the best model (for the best model, that present smaller DIC, this difference is equals zero).

These probability values were obtained for each position of chromosome 7 (Fig. 1).

#### 4. Results and discussion

In summary, the proposed methodology comprising three distinct steps allowed us to conduct a QTL analysis of factors that represent groups of variables, thus enabling the use of a univariate method (Haley et al., 1994) that avoided the involvement of complex multivariate models.

Implementation of the first step of procedure A provided a set of multivariate data without the influence of fixed effects (sex and batch) and random polygenic effect, while for the procedure B the influence of this late effect was not considered. In general, the Gibbs algorithm as implemented in the *GIBBSF90* software (Misztal et al., 2002) converged for all of the parameters, for both procedures, when 50,000 resulting iterations were used. This convergence was verified by the Geweke criterion using the package *BOA* (Bayesian Output Analysis) in the R software.

In the second step, estimates (posterior means) of the factorial loadings  $\lambda_{jk}$  indicated the common factors that accounted for certain groups of traits. Only three common factors (Table 1) explained, respectively to procedures A and B, 85.64% and 80.44% of the total variation from ten original traits. For both procedures, we can note that the fat-related traits (BCD, L, SBT, LR and LL) were highly correlated, and therefore, all of these traits were represented by the first common factor. Similarly, the two loin traits (LEA and LD) could be represented by the second factor; furthermore, the three carcass characteristics (MBCC, MLC and CY) were associated with the third factor. Although the results of the trait classification with respect to each common factor has been the same for the procedures A and B, the differences between traits factorial loadings within each common factor were more evident for the A, i.e., the removal of polygenic effect in the first step provided a better discrimination of traits within each common factor.

In the third step, the estimates (posterior means) of the factorial scores for each animal ( $\hat{c}_i$ ) were subjected to a Bayesian linear regression analysis by considering the model proposed by Haley et al. (1994). Thus, the null regression model (model I), the additive QTL effects model (model II), and the additive and dominance QTL effects model (model III) were compared according to the model posterior probability of each of the 11 positions evaluated (Fig. 1) considering the procedures A and B. Figs. 2a and b shows, respectively for the procedures A and B, the behavior of the probabilities of each of the three models over the SSC 7 positions by considering the regression analysis determined for the first factor, which represent the traits BCD, L, SBT, LR and LL. For the second and third factors, these probabilities did not suggest an additive QTL effect for neither one procedure, and therefore, the graphics corresponding to these factors have been omitted from the figure because they are irrelevant.

Fig. 2 shows that the model II (additive QTL effects) demonstrated the highest posterior probability at position

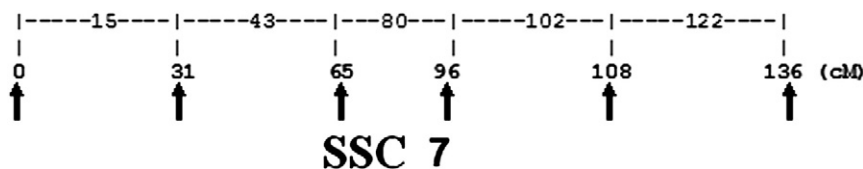


Fig. 1. Schematic illustration of the positions to which the models were compared.

**Table 1**

Estimates (posterior means) of the factorial loadings for each common factor considering the two proposed procedures (A and B).

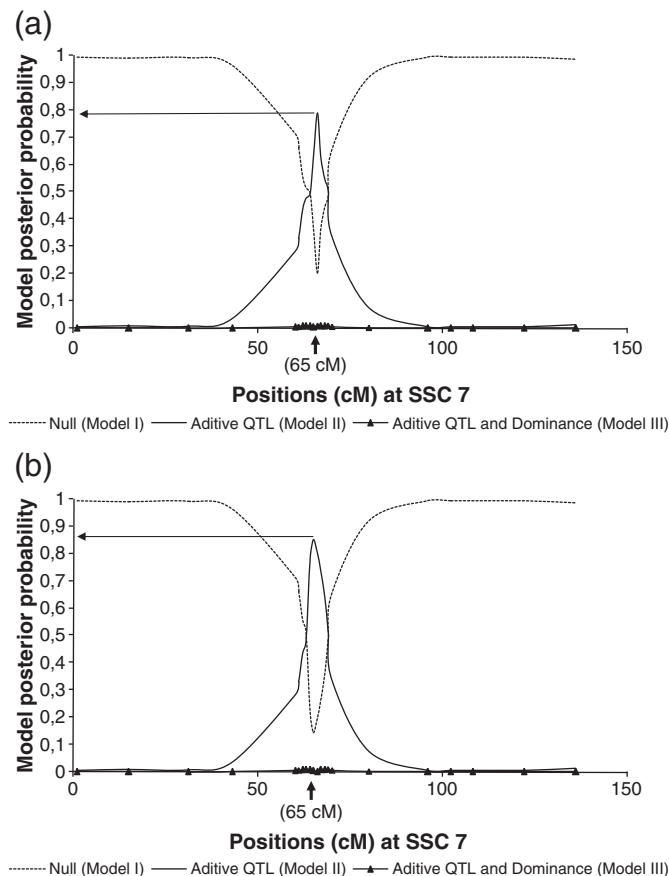
Traits	Common factor (procedure A)			Common factor (procedure B)		
	c1	c2	c3	c1	c2	c3
LR	0.6106	-0.1916	-0.0582	0.4893	-0.1160	-0.0311
LL	0.6028	-0.1186	-0.1179	0.4719	-0.0977	-0.0723
L	0.6185	-0.0781	-0.0926	0.4801	-0.0884	-0.0466
BCD	0.5556	0.0065	-0.1773	0.4444	0.0123	-0.0833
LEA	-0.1273	-0.6553	-0.2388	-0.1011	-0.1553	-0.0711
MBCC	0.3364	-0.4338	0.3000	0.3834	-0.1234	0.1012
MLC	0.3300	-0.3786	0.2801	0.3634	-0.0766	0.0967
SBT	0.5435	-0.2080	0.0577	0.4321	-0.1021	0.0033
LD	0.0254	-0.6876	-0.2438	0.0711	-0.1799	-0.0384
CY	0.2713	-0.3600	0.2608	0.3476	-0.1100	0.0907

65 cM for procedure A (probability of 0.78) and B (probability of 0.86). For the procedure A in this same position, the models I and III, demonstrated probabilities of 0.20 and 0.02, respectively. For the procedure B these values were, respectively, 0.13 and 0.01. In addition, the posterior probabilities of the null model were much higher than other two models for all of the other evaluated positions. In general, the finding that the model with an additive QTL effect presented the higher probability at position 65 cM did

not imply directly the presence of a QTL in this position. The only assumption that can be made is that this position has a 78% and 86% probability, respectively to procedures A and B, of a QTL that simultaneously affects the fat-related traits (BCD, L, SBT, LR and LL). Anyway, we can note that procedure B was more informative about the QTL presence, showing so that a percentage of variance explained by the QTL was removed jointly with polygenic effect (first step) by the use of procedure A.

In order to verify the importance of the position 65 cM for each analyzed trait separately, simple regression analyses were carried out using Qxpak software (Pérez-Enciso and Misztal, 2004). These results showed only significant QTL ( $P < 0.05$ ) for LR and LL at position 65 cM, indicating that the idea of to use common factors rather than the original traits possibly can to increase the QTL detection power for related traits.

Several other studies have investigated traits related to fat deposition and observed significant QTL effects at positions close to 65 cM on SSC 7. Wang et al. (1998) analyzed data obtained from F2 individuals of five different families that were generated via the crossbreeding of two Chinese breeds (Meishan and Minzhu) and three American breeds (Landrace, Duroc, and Hampshire). In that study, a significant QTL was observed at 66 cM for backfat thickness after the first rib. Milan et al. (2002) employed different designs (line-cross and half



**Fig. 2.** Profile of the posterior model probabilities for each regression model considered in the SSC 7 analysis respectively to the procedures A (a) and B (b).

full-sib analyses) to study QTL effects in Meishan  $\times$  Large White animals, and they discovered a significant QTL at position 65 cM for backfat thickness between the third and fourth rib. Yue et al. (2003) evaluated an F2 population with a high degree of heterozygosity that was generated by crossing Meishan, Pietrain, and European wild purebreds, and they detected significant QTLs at 66 and 67 cM for backfat thickness after the last rib and abdominal fat weight, respectively. Demeure et al. (2005) used the half full-sib design to infer significant QTL effects in Large White  $\times$  Meishan animals, and they found significant evidence for a QTL at 69 cM for lower backfat thickness between the third and fourth rib. Sanchez et al. (2006) used backcrossed populations that were established by crossing Large White  $\times$  Meishan animals and found evidence for significant QTL effects at 67 cM for backfat thickness between the third and fourth rib.

With respect to a full Bayesian analysis instead of three step approach, one possible and relatively simple proposal may be the implementation of factor analysis for structuring covariance matrices in a mixed model (de los Campos and Gianola, 2007) with additive genetic and genotypic QTL random effects, being the covariance matrix associated to this late effect a IBD matrix that can be obtained by the Qxpak software (Pérez-Enciso and Misztal, 2004). Thus, the inference about the QTL presence at each chromosome position may be easily realized by the bayesian comparison between complete mixed model and a null model without QTL effect.

## 5. Conclusions

A Three-step Bayesian factor analysis allowed us to calculate the probability of QTLs that simultaneously affect a group of carcass traits for each position of SSC 7. The removal of systematic effects in the first step of the evaluation (procedure B) allowed that the factor analysis, which was performed in the second step, identify three distinct factors that explained 85% of the total traits variation. For the common factor that represented fat-related traits (bacon depth, midline lower backfat thickness, higher backfat thickness on the shoulder; midline backfat thickness after the last rib; midline backfat thickness on the last lumbar vertebrae) the third step of the analysis showed that the highest probability of an additive QTL effect at the 65 cM position was 86%.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.livsci.2011.07.012.

## References

- Band, G.O., Guimarães, S.E.F., Lopes, P.S., Peixoto, J.O., Faria, D.A., Pires, A.V., Figueiredo, F.C., Nascimento, C.S., Gomide, L.A.M., 2005. Relationship between the Porcine Stress Syndrome gene and carcass and performance traits in F2 pigs resulting from divergent crosses. *Genet. Mol. Biol.* 28, 92–96.
- Demeure, O., Sanchez, M.P., Riquet, J., Iannuccelli, N., Demars, J., Fève, K., Kernaleguen, L., Gogué, J., Billon, Y., Caritez, J.C., Milan, D., Bidanel, J.P., 2005. Exclusion of the swine leukocyte antigens as candidate region and reduction of the position interval for the Sus scrofa chromosome 7 QTL affecting growth and fatness. *J. Anim. Sci.* 83, 1979–1987.
- de los Campos, G., Gianola, D., 2007. Factor analysis models for structuring covariance matrices of additive genetic effects: a Bayesian implementation. *Genet. Sel. Evol.* 39, 481–494.
- Faria, D.A., Guimarães, S.E.F., Lopes, P.S., Pires, A.V., Paiva, S.R., Sollero, B.P., Wenceslau, A.A., 2006. Association between G316A growth hormone polymorphism and economic traits in pigs. *Genet. Mol. Biol.* 29, 634–640.
- 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.
- Henderson, C.R., 1984. Applications of linear models in animal breeding. University of Guelph, Guelph. 462 p.
- Johnson, R., Wichern, D., 2007. Applied multivariate statistical analysis, Sixth edition. Prentice-Hall, London. 816 p.
- Lander, E.S., Green, P., 1987. Construction of multi-locus genetic linkage maps in humans. *Proc. Natl. Acad. Sci.* 84, 2363–2367.
- Meyer, K., 2009. Factor-analytic models for genotype  $\times$  environment type problems and structured covariance matrices. *Genet. Sel. Evol.* 41, 21.
- Milan, D., Bidanel, J.P., Iannuccelli, N., Riquet, J., Amigues, Y., Grund, J., Le Roy, P., Renard, C., Chevalet, C., 2002. Detection of quantitative trait loci for carcass composition traits in pigs. *Genet. Sel. Evol.* 34, 705–728.
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D.H., 2002. BLUPF90 and related programs. World Congress on Genetics Applied to Livestock Production, Montpellier. Proceedings WCGALP, 28,07.
- Newton, M., Raftery, A., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. R. Stat. Soc. Series B* 56, 3–48.
- Pérez-Enciso, M., Misztal, I., 2004. Qxpak: a versatile mixed model application for genetical genomics and QTL analyses. *Bioinformatics* 20, 2792–2798.
- R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 3-900051-07-0, <http://www.R-project.org>.
- Sanchez, M.P., Riquet, J., Iannuccelli, N., Gogue, J., Billon, Y., Demeure, O., Caritez, J.C., Burgaud, G., Fève, K., BONNET, M., PERY, C., 2006. Effects of quantitative trait loci on chromosomes 2, 4, and 7 on growth carcass, and meat quality traits in backcross Meishan  $\times$  Large White pigs. *J. Anim. Sci.* 84, 526–537.
- Stearns, T.M., Beaver, J.E., Southey, B.R., Ellis, M., Mckeith, F.K., Rodriguez-Zas, S.L., 2005. Evaluation of approaches to detect quantitative trait loci for growth, carcass, and meat quality on swine chromosomes 2, 6, 13, and 18, II. Multivariate and principal component analyses. *J. Anim. Sci.* 83, 2471–2481.
- Szydlowski, M., Buszka, A., Mackowski, M., Lechniak, D., Switonski, M., 2011. Polymorphism of genes encoding cytokines IL6 and TNF is associated with pig fatness. *Livestock Science* 136, 150–156.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* 64, 583–639.
- Wang, L., Yu, T.P., Tuggle, H., Liu, C., Rothschild, M.F., 1998. A direct search for quantitative trait loci on chromosomes 4 and 7 in pigs. *J. Anim. Sci.* 76, 2560–2567.
- Wilberg, M.J., Bence, J.R., 2008. Performance of deviance information criterion model selection in statistical catch-at-age analysis. *Fisheries Research* 93, 212–221.
- Yue, G.B., Stratil, A., Cepica, S., Schroffler, J., Schrofflova, D., Fontanesi, L., Cagnazzo, M., Moser, G., Bartenschlager, H., Reiner, G., 2003. Linkage and QTL mapping for Sus scrofa chromosome 7. *J. Anim. Breed. Genet.* 120, 56–65.