

A BAYES FACTOR APPROACH TO DISCRIMINATE BETWEEN LINKED OR PLEIOTROPIC QTL

L. Varona¹, L. Gómez-Raya¹, W. M. Rauw¹, A. Clop², C. Ovilo³ and J. L. Noguera¹

¹Àrea de Producció Animal. Centre UdL-IRTA. Lleida, 25198, Spain.

²Departament de Ciència Animal i dels Aliments. UAB. Bellaterra, 08193, Spain

³Departamento de Mejora Genética Animal. CIT-INIA. Madrid, 28040, Spain

INTRODUCTION

Most of the experiments to map QTLs involve recording of several traits, often genetically correlated between them. Several authors have tried to solve the problem of discriminating between linked and pleiotropic QTLs (Cheverud *et al.*, 1997 ; Almasy *et al.*, 1997 ; Lebreton *et al.*, 1998 ; Knott and Haley, 2000). From a bayesian point of view, the comparison between alternative models is solved by the calculation of Bayes Factors (Kass and Raftery, 1995). The Bayes Factor is the ratio between the marginal probabilities of data under both models after integrating out all the parameters. The objective of this paper is to propose a Bayes Factor approach for testing between linked and pleiotropic QTL. First, we present the general Bayes Factor, and, secondly, two particular examples from an experimental F₂ cross between outbred lines of pigs are presented.

METHOD

Here, we describe a general Bayes Factor approach for comparing bivariate models with linked or pleiotropic QTLs. In the linked QTL model, bivariate data (y_1, y_2) are described by a probability function conditioned to a set of parameters for both traits (θ_1, θ_2) and locations for the QTL in both traits (λ_1, λ_2):

$$y_1, y_2 | \theta_1, \theta_2, \lambda_1, \lambda_2$$

In contrast, in the pleiotropic QTL model, data from both traits are described by a probability function that includes only one QTL location (λ_p) and the same set of parameters of the previous model.

$$y_1, y_2 | \theta_1, \theta_2, \lambda_p$$

Then, the Bayes Factor is calculated from the ratio of marginal probabilities of data under both models:

$$BF = \frac{p_l(y_1, y_2)}{p_p(y_1, y_2)} = \frac{\frac{p_l(y_1, y_2 | \theta_1, \theta_2, \lambda_1, \lambda_2) p_l(\theta_1, \theta_2) p_l(\lambda_1, \lambda_2)}{p_l(\theta_1, \theta_2, \lambda_1, \lambda_2 | y_1, y_2)}}{\frac{p_p(y_1, y_2 | \theta_1, \theta_2, \lambda_p) p_p(\theta_1, \theta_2) p_p(\lambda_p)}{p_p(\theta_1, \theta_2, \lambda_p | y_1, y_2)}}$$

If we assume that

$$p_l(\theta_1, \theta_2) = p_p(\theta_1, \theta_2),$$

and

$$p_l(\mathbf{y}_1, \mathbf{y}_2 | \theta_1, \theta_2, \lambda_1 = \lambda_2 = k_1) = p_p(\mathbf{y}_1, \mathbf{y}_2 | \theta_1, \theta_2, \lambda_p = k_1)$$

when the locations for both QTL are the same, then:

$$BF = \frac{\frac{p_l(\lambda_1 = \lambda_2 = k_1)}{p_l(\theta_1, \theta_2, \lambda_1 = \lambda_2 = k_1 | \mathbf{y}_1, \mathbf{y}_2)}}{\frac{p_p(\lambda_p = k_1)}{p_p(\theta_1, \theta_2, \lambda_p = k_1 | \mathbf{y}_1, \mathbf{y}_2)}} = \frac{\frac{p_l(\lambda_1 = \lambda_2 = k_1)}{p_l(\theta_1, \theta_2 | \mathbf{y}_1, \mathbf{y}_2, \lambda_1 = \lambda_2 = k_1)} p_l(\lambda_1 = \lambda_2 = k_1 | \mathbf{y}_1, \mathbf{y}_2)}{\frac{p_p(\lambda_p = k_1)}{p_p(\theta_1, \theta_2 | \mathbf{y}_1, \mathbf{y}_2, \lambda_p = k_1)} p_p(\lambda_p = k_1 | \mathbf{y}_1, \mathbf{y}_2)}}$$

As $p_l(\theta_1, \theta_2 | \mathbf{y}_1, \mathbf{y}_2, \lambda_1 = \lambda_2 = k_1) = p_p(\theta_1, \theta_2 | \mathbf{y}_1, \mathbf{y}_2, \lambda_p = k_1)$, then:

$$BF = \frac{p_l(\lambda_1 = \lambda_2 = k_1) p_p(\lambda_p = k_1 | \mathbf{y}_1, \mathbf{y}_2)}{p_l(\lambda_1 = \lambda_2 = k_1 | \mathbf{y}_1, \mathbf{y}_2) p_p(\lambda_p = k_1)}$$

From that, the posterior probability of linkage model is $BF/(1+BF)$ and, the posterior probability of pleiotropy model $1/(1+BF)$.

The procedure has been implemented using a MCMC algorithm.

DATA

As an example, we used data from an F2 experiment between Landrace and Iberian pigs. The experiment is described by Pérez-Enciso *et al.* (2000) and Ovilo *et al.* (2000). The pedigree was consisted of 3 Iberian boars, 31 Landrace sows, 6 F1 boars, 73 F1 sows and 321 F2 individuals, from 58 full sib families. Among all the traits recorded, we selected the following traits: L* measure of color by Minolta (L*), Backfat depth between 3rd and 4th rib (DFAT) and % of linoleic acid (LIN). In the fourth chromosome, all individuals were genotyped for the SW2404, S0301, S0001, SW839, DECR2, S0214, SW445, S0097 located at positions 0.0, 40.8, 59.5, 72.8, 78.8, 95.0, 116.8 and 134.4 cM, respectively. Genetic mapping was performed using the CRI-MAP using the BUILD option (Green *et al.*, 1990). The selected traits were significant for a QTL in the SSC4 using regression procedures (Haley *et al.*, 1994). L* presents a maximum F value at location 109, DFAT at 71 and LIN at 75 cM.

CASE I. BAYES FACTOR ANALYSIS BETWEEN L* AND % LINOLEIC ACID

For the pleiotropy model, the likelihood was:

$$p(\mathbf{y}_1, \mathbf{y}_2 | \beta_1, \beta_2, a_1, a_2, d_1, d_2, \lambda_p, \sigma_{e1}^2, \sigma_{e2}^2) = N \left(\begin{array}{c} \mathbf{X}\beta_1 + \mathbf{p}_1(\lambda_p)a_1 + \mathbf{p}_2(\lambda_p)d_1 \\ \mathbf{X}\beta_2 + \mathbf{p}_1(\lambda_p)a_2 + \mathbf{p}_2(\lambda_p)d_2 \end{array}, \begin{array}{ccc} \sigma_{e1}^2 & 0 & \\ 0 & \sigma_{e2}^2 & \end{array} \right)$$

Where \mathbf{X} is the incidence matrix, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the vectors for systematic effects (sex and family) for traits 1 and 2, respectively, a_1 and a_2 are the additive values and d_1 and d_2 are the dominance values of the QTLs for both traits, λ_p is the location of the QTLs, $\mathbf{p}_1(\lambda)$ is the vector of differences between the probability of being homozygous for the Iberian allele and the probability of being homozygous for the Landrace allele at location λ , $\mathbf{p}_2(\lambda)$ is the vector of the probabilities of being heterozygous. Finally, σ_{e1}^2 and σ_{e2}^2 are the residual variances for both traits. Prior densities were assumed flat for all the parameters.

For the linkage model, the likelihood was:

$$p(\mathbf{y}_1, \mathbf{y}_2 | a_1, a_2, d_1, d_2, \lambda_1, \lambda_2, \sigma_{e1}^2, \sigma_{e2}^2) = N \left(\begin{array}{cc} \mathbf{X}\boldsymbol{\beta}_1 + \mathbf{p}_1(\lambda_1)a_1 + \mathbf{p}_2(\lambda_1)d_1 & \sigma_{e1}^2 & 0 \\ \mathbf{X}\boldsymbol{\beta}_2 + \mathbf{p}_1(\lambda_2)a_2 + \mathbf{p}_2(\lambda_2)d_2 & 0 & \sigma_{e2}^2 \end{array} \right)$$

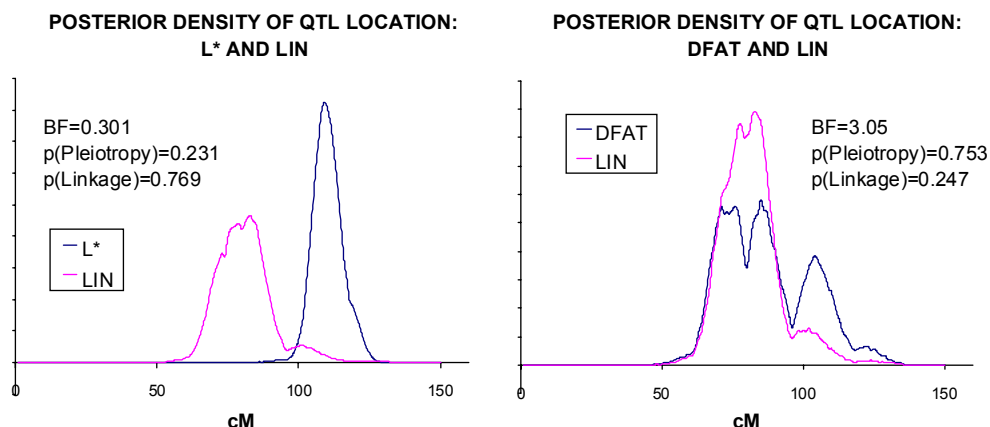
where λ_1 and λ_2 are the locations of the QTLs for both traits. As in the pleiotropy model, prior distributions of all parameters were assumed to be flat. In both models, the location of the molecular markers along the SSC4 was assumed to be known.

The Bayes Factor between the pleiotropy and the linkage model was 0.301. Thus, the posterior probability of the linkage model was 0.769 and the posterior probability of the pleiotropy model was 0.231. The marginal posterior distributions for the location of the QTLs for both traits are presented in figure I. Posterior mean estimates of the percentage of variance explained by the QTL was 13.4% and 8.4% with posterior standard deviations of 4.2% and 3.5% for L* and LIN, respectively. Therefore, QTLs affecting L* and LIN are linked with higher probability, after using the information provided by the data.

CASE 2. BAYES FACTOR ANALYSIS BETWEEN BACKFAT DEPTH AND % LINOLENIC ACID

The model of analysis was the same as in the Case 1. Here, the Bayes Factor between the pleiotropy and the linkage model was 3.05. Thus, the posterior probability for linkage or pleiotropy model were 0.247 and 0.753. Marginal posterior distribution for location of the QTLs are presented in Figure I. Posterior mean estimates for the percentage of variance explained by the QTL were 7.9% and 8.4% with posterior standard deviations of 3.4% and 3.5% for DFAT and LIN, respectively. Therefore, the effects of the QTLs affecting DFAT and LIN are pleiotropic with higher probability, after using the information provided by the data.

Figure 1.



CONCLUSION

The proposed procedure provides a practical tool to discriminate between pleiotropy and linkage when QTLs are detected for two traits. The method is easily implemented to any model that includes the location of the QTLs as a parameter of the model. An important advantage of this procedure is that it is not necessary to establish null or alternative hypotheses, because it provides directly the posterior probability of both models. The Bayes Factor is the ratio of the probabilities of the same location for the QTLs in both traits before and after the introduction of the information provided by the data. However, it be noted that when the two QTLs are very closely linked there is no way to discriminate between linkage and pleiotropy. Actually, the procedure discriminate between two situations: 1) the QTLs are at the same location and 2) the QTLs are a different location.

ACKNOWLEDGMENTS

We want to acknowledge the CICYT grants AGF96-2510 and AGF99-0284. We are specially grateful to Pere Borràs and Eva Ramells together with all the personnel in Nova Genètica S.A.

REFERENCES.

- Almansy *et al.* (1997) *Genet. Epidemiol.* **14** : 953-958.
- Cheverud *et al.* (1997) *Evolution* **51** : 2006-2016.
- Green *et al.* (1990) Crimap documentation.
- Haley *et al.* (1994) *Genetics* **136** : 1195-1207.
- Kass and Raftery (1995) *J. Am. Stat. Assoc* **90** : 773-795.
- Knott and Haley (2000) *Genetics* **156** : 899-911.
- Lebreton, C.H. *et al.* (1998) *Genetics* **150** : 931-943.
- Ovilo *et al.* (2000) *Mamm. Genome* **11** : 344-346.
- Pérez-Enciso *et al.* (2000) *J. Anim. Sci* **78** : 2525-2531.