

Complex Segregation Analysis of Categorical Traits in Farm Animals: Comparison of Linear and Threshold Models

H. N. Kadarmideen* and H. Ilahi

Statistical Animal Genetics Group, Institute of Animal Science, Swiss Federal Institute of Technology
ETH Zentrum, CH-8092 Zuerich, Switzerland

ABSTRACT : Main objectives of this study were to investigate accuracy, bias and power of linear and threshold model segregation analysis methods for detection of major genes in categorical traits in farm animals. Maximum Likelihood Linear Model (MLLM), Bayesian Linear Model (BALM) and Bayesian Threshold Model (BATM) were applied to simulated data on normal, categorical and binary scales as well as to disease data in pigs. Simulated data on the underlying normally distributed liability (NDL) were used to create categorical and binary data. MLLM method was applied to data on all scales (Normal, categorical and binary) and BATM method was developed and applied only to binary data. The MLLM analyses underestimated parameters for binary as well as categorical traits compared to normal traits, with the bias being very severe for binary traits. The accuracy of major gene and polygene parameter estimates was also very low for binary data compared with those for categorical data: the later gave results similar to normal data. When disease incidence (on binary scale) is close to 50%, segregation analysis has more accuracy and lesser bias, compared to diseases with rare incidences. NDL data were always better than categorical data. Under the MLLM method, the test statistics for categorical and binary data were consistently unusually very high (while the opposite is expected due to loss of information in categorical data), indicating high false discovery rates of major genes if linear models are applied to categorical traits. With Bayesian segregation analysis, 95% highest probability density regions of major gene variances were checked if they included the value of zero (boundary parameter); by nature of this difference between likelihood and Bayesian approaches, the Bayesian methods are likely to be more reliable for categorical data. The BATM segregation analysis of binary data also showed a significant advantage over MLLM in terms of higher accuracy. Based on the results, threshold models are recommended when the trait distributions are discontinuous. Further, segregation analysis could be used in an initial scan of the data for evidence of major genes before embarking on molecular genome mapping. (*Asian-Aust. J. Anim. Sci.* 2005, Vol 18, No. 8 : 1088-1097)

Key Words : Major Gene, Segregation Analysis, Linear and Threshold Models, Maximum Likelihood, Bayesian Method, Pigs

INTRODUCTION

Traditional quantitative genetics theory and its application to animal breeding are based on the classical assumption that traits are controlled by a very large number of independent genes each having a small effect. Nevertheless, several genes having major effects on quantitative traits of economic importance have been identified in livestock. Notable examples are double muscling gene in cattle (Hanset and Michaux, 1985), Booroola gene in sheep (Piper and Bindon, 1982), the gene determining high milk flow in goats (Ricordeau et al., 1990) and Natural resistance-associated macrophage protein 1 (NRAMP1) gene controlling resistance to various intracellular parasites in pigs (Yan et al., 2004).

Segregation analysis (Elston and Stewart, 1971) or Mixed Inheritance Models or MIM (to denote polygenic and major gene influence on the trait) has been proposed as a suitable and powerful method to identify segregating single major gene in livestock populations (Le Roy et al.,

1989; Hill and Knott, 1990; Knott et al., 1991). It involves maximizing and comparing the likelihoods of the data under different genetic transmission models to determine whether the inheritance of the trait is controlled by a major gene. A significant improvement in the likelihood obtained by incorporating a major gene in the model provides evidence for the segregation of major gene in the population under study (Knott et al., 1991). In a Bayesian inference framework, the Gibbs sampler algorithm was adapted by Guo and Thompson (1994) in order to solve computing problems in complex pedigrees in animal genetics. The Gibbs sampling algorithms have now found a wide-spread use in genetic analysis of quantitative traits recorded in pedigreed animal populations, due to its flexibility in solving complex and demanding statistical models, especially for categorical traits (e.g. Kadarmideen et al., 2001; Lee, 2002).

In segregation analysis of pedigreed animal populations, both maximum likelihood and Bayesian approaches were first developed for normally distributed traits. For example, Bayesian models with Gibbs sampling has been recently applied to dairy cows in Switzerland to detect major genes for milking speed (Ilahi and Kadarmideen, 2004) or egg weight in chickens (Hagger et al., 2004), a normally

* Corresponding Author: Haja N. Kadarmideen. Tel: +41-(0)1-632-3266, Fax: +41-(0)1-632-1260, E-mail: haja.kadarmideen@inw.agr.ethz.ch

Received November 1, 2004; Accepted March 7, 2005

distributed continuous trait. However, in many animal breeding applications, the data comprises observations which are expressed in two or more categories, representing binary and categorical traits, respectively (e.g. diseases traits, ovulation rate in sheep, non-return rates in cows, degree of calving difficulty, conformation and type traits). Linear and non-linear (threshold) models have been applied and compared using disease data from dairy cattle (e.g. Kadarmideen et al., 2000a), in quantitative trait loci (QTL) mapping in livestock populations and their efficiency (Kadarmideen et al., 2000b; Yi and Xu, 2000; Kadarmideen and Dekkers, 2001). However, no study has so far evaluated and compared such models (linear and threshold model) or the statistical methods (Maximum likelihood versus Bayesian method based on Monte carlo Markov Chain algorithms) in the context of segregation analysis of binary or categorical traits in animal populations.

The main objectives of this study were to compare linear models (LM) and threshold models (TM) segregation analyses based on maximum likelihood and Bayesian methods, respectively.

i. Investigate the impact of distribution of the trait (normal versus categorical or binary data) on the accuracy and power of detecting major genes in the population, using maximum likelihood linear model (MLLM) method.

ii. Investigate the impact of different incidences of binary trait on the accuracy and power of detection of major genes by segregation analysis under both MLLM and Bayesian Threshold Model (BATM) method.

iii. Apply the developed Bayesian methods (under linear and threshold models) to osteochondral diseases (categorical scores) in Swiss pig populations.

MATERIALS AND METHODS

Liability theory

The genetic analyses for categorical traits are difficult because the observed phenotype cannot be described by a linear function of genetic and environmental effects. Wright (1934) proposed the assumption that there is a continuous underlying variable, called *liability*, which determines the expression of a categorical trait. The link between the observable discrete variable and the underlying variable is generated by a set of fixed thresholds. The underlying liability variable is then described by the usual linear model (Gianola, 1982; Gianola and Foully, 1983; Falconer and Mackay, 1996). These liability-threshold models have been applied to gene or quantitative trait loci (QTL) detection and mapping (Xu and Atchley, 1996; Yi and Xu, 2000; Kadarmideen et al., 2000b; Kadarmideen and Dekkers, 2001; Kadarmideen and Janss, 2003).

Data simulations

Normally distributed liability (NDL) data : The liability data were simulated using a mixed inheritance model (polygenes plus a major gene) and according to a hierarchical and balanced family structure: one population consisted of 20 sire families with 100 dams per sire, which resulted in 2000 dams. A total of 3 records/phenotypes per dam was simulated (i.e. 300 records per sire). Therefore the total number of records in the population is 6000. We assumed more than one phenotype per dam, because in most cases, livestock data sets consist of repeated records. Sires and dams were assumed to be unrelated and have the same allele frequency of the major gene.

The liability data were simulated as follows:

$$z_{ij} = m_i + u_{ij} + pe_{ij} - e_{ij} \quad (1)$$

Where z_{ij} is the liability, m_i is the effect of the i^{th} genotype at the major gene, u_{ij} is the polygenic effect of the j^{th} individual bearing the i^{th} genotype, with $u_{ij} \sim N(0, \sigma_u^2)$, pe_{ij} is the permanent environmental effect, with $pe_{ij} \sim N(0, \sigma_{pe}^2)$ and e_{ij} is the residual effect, with $e_{ij} \sim N(0, \sigma_e^2)$, where σ_u^2 , σ_{pe}^2 and σ_e^2 are genetic, permanent environmental and residual variances, respectively. The single major gene is assumed to be an additive, biallelic (A_1 and A_2) and an autosomal locus with Mendelian transmission probabilities. We consider here that $p_1 = 0.6$ and $p_2 = 1 - p_1$ are the frequencies of alleles A_1 and A_2 . Three genotypes can be encountered: A_1A_1 , A_1A_2 and A_2A_2 , with a frequency of p_1^2 , $2p_1p_2$ and p_2^2 , respectively. The A_2 allele is assumed to increase the trait value, and is called the favorable allele (depending on whether the increase in trait value is favorable or not). Further, we assume no dominance, the difference between the values of the two homozygotes (a) was 3.7 phenotypic standard deviation units of the trait, and the major gene variance, $\sigma_m^2 = 2p_1p_2a^2$ (Falconer and Mackay, 1996) was 5.88, which is 80% of the total genetic variance. Some traits in animal populations that are controlled by major gene effects close to this magnitude have been reported in literature (Ilahi et al., 2000; Hagger et al., 2004).

The liability data were simulated using heritability, h^2 , of 0.41 and repeatability, r , of 0.52 on the liability scale (the total heritability and repeatability taking into account the major gene effect were 0.78 and 0.82 respectively). The genotype of the offspring was determined according to the Mendelian transmission probabilities. The polygenic effect of the offspring was determined as the summation of the mean of the parents' polygenic effect and the Mendelian sampling effect. The true values of parameters (major gene and polygene) used in the simulation of the population are illustrated in the Table 1.

Categorical data : The simulated NDL data, z_{ij} , were standardized using the average, μ and the standard deviation,

Table 1. True values of parameters and parameter estimates by maximum likelihood linear model segregation analyses for liability and categorical data: means and standard deviations of 100 replicates

| Parameters | True values | Liability data | | Categorical data | |
|-----------------|-------------|---------------------------------|-----------------------------|---------------------------------|-----------------------------|
| | | Polygenic inheritance (H^0) | Mixed inheritance (H_1) | Polygenic inheritance (H^0) | Mixed inheritance (H_1) |
| μ | 0 | -0.10**±0.28 | - | 2.05**±0.15 | - |
| $\mu_{A_1A_1}$ | -3.50 | - | -3.47±0.20 | - | 1.51**±0.09 |
| $\mu_{A_1A_2}$ | 0 | - | 0.04±0.18 | - | 1.58**±0.07 |
| $\mu_{A_2A_2}$ | 3.50 | - | 3.55±0.19 | - | 4.05**±0.09 |
| $f(A_1A_1)$ | 0.36 | - | 0.35±0.05 | - | 0.39±0.10 |
| $f(A_1A_2)$ | 0.48 | - | 0.48±0.09 | - | 0.48±0.13 |
| $f(A_2A_2)$ | 0.16 | - | 0.17±0.03 | - | 0.13±0.04 |
| σ_u^2 | 1.44 | 4.74**±0.77 | 1.30**±0.26 | 0.65**±0.13 | 0.17**±0.04 |
| σ_{pe}^2 | 0.36 | 2.38**±0.66 | 0.45**±0.18 | 0.59**±0.12 | 0.07**±0.03 |
| σ_e^2 | 1.69 | 1.68±0.03 | 1.68±0.03 | 0.43**±0.02 | 0.43**±0.02 |
| σ_m^2 | 5.88 | - | 3.45**±0.80 | - | 0.48**±0.14 |
| Heritability | 0.41 | 0.54**±0.07 | 0.38±0.06 | 0.39±0.07 | 0.25**±0.06 |
| Repeatability | 0.52 | 0.80**±0.01 | 0.51±0.02 | 0.74**±0.01 | 0.36**±0.03 |

Means with ** are significantly different from the true value ($p < 0.01$).

z_y of the trait as: $z_y^* = (z_y - \mu) / \sigma_x$, and hence z_y^* is the standardized liability data with $z_y^* \sim N(0, 1)$. Then the standardized liability data was transformed into categorical data (5 categories) using 4 different thresholds. The expression of the discrete phenotype is based on the existence of thresholds t_λ ($\lambda = 1, \dots, 4$): when the standardized NDL data z_y^* is in the range $(t_{\lambda-1}, t_\lambda)$, the discrete phenotype $y_y^c = \lambda$. The extreme thresholds t_1 and t_4 are respectively $-\infty$ and $+\infty$. Category threshold and their corresponding incidences used in this study were: $t_1 = 0$, $t_2 = 0.674$, $t_3 = 1.036$ and $t_4 = 1.282$, corresponding to 4 category incidences. $p_1 = 50\%$, $p_2 = 25\%$, $p_3 = 15\%$ and $p_4 = 10\%$.

Binary (0:1) data : Binary data are the simplest form of categorical traits with just two categories. Based on the liability concepts, the standardized liability data z_y^* could be transformed into binary data as follows:

$$\begin{cases} \text{if } z_y^* \geq t \text{ then } y_y^b = 1 \\ \text{if } z_y^* < t \text{ then } y_y^b = 0 \end{cases}$$

Where t is the threshold point. Here y_y^b taking value of '1' could be considered as diseased and '0' as healthy, thus representing liability model for complex polygenic diseases.

The values for t were chosen in such a way that it represents two scenarios: a less common disease with 15% and more common disease with 40% incidence. The corresponding values of thresholds t were: $t = 1.036$ for 15% and $t = 0.253$ for 40% incidence.

The NDL data, categorical and binary data resulting from the transformation of the same NDL data were kept for segregation analyses by MLLM and BATM methods.

Maximum likelihood linear model (MLLM) analysis

There were 4 types of data sets. The original NDL data,

the categorical data and the two binary data sets with 15% and 40% incidences. Same MLLM method was performed on all datasets. Simulations and analyses were replicated 100 times for each data set. Different values of parameters were used as initial values for the calculations of the estimated parameters. The segregation analysis method used in this study was based on the comparison of the likelihoods under two inheritance hypotheses (Le Roy et al., 1990; Ilahi et al., 2000; Bodin et al., 2002):

Mixed inheritance hypothesis (H_1) : This model describes the genetic transmission of the simulated trait by polygenic effects and a single major gene effect. The parameters to be estimated are: the mean of each genotype ($\mu_{A_1A_1}$, $\mu_{A_1A_2}$, $\mu_{A_2A_2}$), the three variance components (σ_u^2 , σ_{pe}^2 , σ_e^2) and the genotypic frequencies $f(A_1A_1)$, $f(A_1A_2)$ and $f(A_2A_1) = 1 - f(A_1A_1) - f(A_2A_2)$. These estimated parameters allowed the computation of the within major gene heritability, the repeatability and the fraction of total genetic variance explained by the major gene.

Polygenic inheritance hypothesis (H_0) : This model, which is a sub-model of the H_1 mixed inheritance hypothesis, is given by $\mu_{A_1A_1} = \mu_{A_1A_2} = \mu_{A_2A_2} = \mu$. In this case, the parameters to be estimated are: μ , σ_u^2 , σ_{pe}^2 , σ_e^2 from which we can compute the heritability, and the repeatability.

The likelihoods λ_0 and λ_1 were computed respectively for both hypotheses H_0 and H_1 . The test statistic is the likelihood ratio given by $LR = -2 \log(\lambda_0 / \lambda_1)$.

Using the simulated population structure and following the model (1), λ_1 can be written, (Le Roy et al., 1995; Ilahi, 1999) as:

$$\lambda_1 = \prod_{i=1}^{20} \sum_{m_i}^3 P_{m_i} \int f(u_i) f(y_i / u_i, m_i) \times \prod_{j=1}^{200} \sum_{m_j}^3 P_{m_j} \int f(u_j) f(y_j / u_j, m_j) \times \prod_{k=1}^{630} \sum_{m_k}^3 P(m_{pk} | m_i, m_j) f(y_{pk} / u_i, u_j, m_{pk}, p_{pe}) du_i du_j$$

with:

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{1}{2} \frac{u_i^2}{\sigma_u^2}\right)$$

$$f(u_{ij}) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{1}{2} \frac{u_{ij}^2}{\sigma_u^2}\right)$$

$$f(y_i / u_i, m_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2} \frac{(y_i - u_i - \mu_{m_i})^2}{\sigma_e^2}\right)$$

$$f(y_{ij} / u_{ij}, m_{ij}) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2} \frac{(y_{ij} - u_{ij} - \mu_{m_{ij}})^2}{\sigma_e^2}\right)$$

and

$$f(y_{ijk} / u_i, u_{ij}, m_{ik}, p_{e_{ik}}) = \frac{1}{\sqrt{2\pi(\sigma_u^2 - \sigma_e^2 / 2 - \sigma_{pe}^2)}} \exp\left(-\frac{1}{2} \frac{(y_{ijk} - (u_i - u_{ij}) / 2 - \mu_{m_{ik}} - p_{e_{ik}})^2}{\sigma_u^2 + \sigma_e^2 / 2 - \sigma_{pe}^2}\right)$$

and λ_0 was defined as:

$$\lambda_0 = \prod_{i=1}^{30} \int_{-\infty}^{\infty} f(u_i) f(y_i / u_i) \cdot \prod_{j=1}^{300} \int_{-\infty}^{\infty} f(u_{ij}) f(y_{ij} / u_{ij}) \cdot \prod_{k=1}^{600} \int_{-\infty}^{\infty} f(y_{ijk} / u_i, u_{ij}) du_i du_{ij} dy_{ijk}$$

This likelihood ratio is compared to the value of χ_d^2 with degrees of freedom d equal the difference in number of parameters between the mixed and polygenic inheritance hypotheses (Le Roy et al., 1989; Kadarmideen et al., 2000b; Kadarmideen and Dekkers, 2001). In this analysis, $d = 4$.

The estimation of parameters maximising the likelihoods was carried out using the Gauss-Hermit quadrature (D01BAF) and optimization (E04JBF) subroutines of the NAG FORTRAN Library (1990) with a quasi-Newton algorithm in which the derivatives were estimated by finite differences.

Bayesian analysis

The analyses were carried out on the same simulated binary data sets (with 15 and 40% incidences) using a Bayesian threshold model (BATM) with Gibbs sampling. MAGGIC software package (Janss, 1998) was used to estimate the genetic parameters of the population. This method constructs Monte Carlo chains of realizations of the model parameters through Gibbs-sampling. These samples constitute the marginal posterior distributions of the model parameters, from which Bayesian inferences on these parameters can be drawn. The method was adapted to our current study as follows.

Animals starting genotypes were randomly allocated, and the initial gene frequency was 0.50. Uniform prior

distributions were assumed for variance components: polygenic variance σ_u^2 , permanent environmental variance σ_{pe}^2 and residual variance σ_e^2 as well as for additive major gene effect a , and for allele frequencies p_1 and p_2 . Variance components are *a priori* positive, and the allele frequencies are bounded between zero and one including the bounds (Janss et al., 1995; Janss, 1998). The joint density of all the parameters to be estimated, given the data y^c or y^b , is denoted as follows:

$$f(\mu, \sigma_u^2, \sigma_{pe}^2, \sigma_e^2, a, p_1 / y^b) \tag{2}$$

$$f(\mu, \sigma_u^2, \sigma_{pe}^2, \sigma_e^2, a, p_1 / y^c)$$

Multivariate normal distributions were adopted as priors for animal genetic and permanent environmental effects:

$$f(u | \mathbf{A}, \sigma_u^2) \sim N(\mathbf{0}, \mathbf{A} \sigma_u^2)$$

$$f(pe | \sigma_{pe}^2) \sim N(\mathbf{0}, \mathbf{I} \sigma_{pe}^2)$$

$$f(e | \sigma_e^2) \sim N(\mathbf{0}, \mathbf{I} \sigma_e^2)$$

Here, \mathbf{A} is the additive relationship matrix and \mathbf{I} is the identity matrix. For the dispersion parameters, σ_u^2 , σ_{pe}^2 and σ_e^2 , uniform bounded priors were adopted.

In case of Bayesian Threshold Model (BATM) applied to a binary disease, the observed data vector, y^b is replaced by the underlying liability z , with the same model terms as in (1). Further details and distributional assumptions in threshold models are given in Janss (2004).

The Gibbs sampler based on Markov chain is primarily used to generate samples from the joint density (Geman and Geman, 1984; Gelfand and Smith, 1990). These samples allow the study of all marginal densities from the joint density. The Gibbs Markov chain is a continuing series of realisations for the parameters given in the joint density (2). Let $\theta^{[t]} = (\mu^{[t]}, \sigma_u^{2[t]}, \sigma_{pe}^{2[t]}, \sigma_e^{2[t]}, a^{[t]}, p_1)$ denote the set of realisations for the parameters to be estimated at cycle t in the Gibbs chain. Construction of Gibbs chain requires a set of realisations $\theta^{[t+1]}$, given the current set of realisations $\theta^{[t]}$. The Gibbs sampler is generally used to study marginal posteriors densities of parameters, considering other parameters as nuisances. The marginal distributions for each parameter are updated in the order given in the joint density (2):

General mean : The construction of cycle $t+1$ from t of the general mean μ is:

$$f(\mu | \sigma_u^2, \sigma_{pe}^2, \sigma_e^2, a, p_1) \sim (y_{ij} - m^{[t]} - u^{[t]} - pe^{[t]}) \tag{3}$$

Polygenic, permanent environmental and residual

variance : Variance components follow inverted chi-square distributions (Janss et al., 1995). the new realisations for σ_u^2 , σ_{pe}^2 and σ_e^2 are obtained as:

$$\text{Sample } \sigma_u^2 \text{ as } \mu^{[t+1]}, A^{-1}u^{[t-1]}, \chi^2(q-2) \quad (4)$$

$$\text{Sample } \sigma_{pe}^2 \text{ as } \mu^{[t+1]}, A^{-1}u^{[t-1]}, \chi^2(q-2) \quad (5)$$

$$\text{Sample } \sigma_e^2 \text{ as } e^{[t-1]}, e^{[t-1]}, \chi^2(n-2) \quad (6)$$

where A is the numerator relation matrix, q is number of animals, n is the total number of observations, and $\chi^2(q-2)$ and $\chi^2(n-2)$ are random deviates from chi-square distributions with q-2 and n-2 degrees of freedom.

Major gene effect : Using genotypes as a known classification factor, effect of a major gene effect is estimated as the deviation of homozygotes from an assumed mean of zero (Janss et al., 1995).

$$\text{Sample } a^{[t-1]} \text{ from } N((\bar{y}_3 - \bar{y}_1)/(n_1 + n_3), \sigma_e^{2(t+1)/(n_1 + n_3)}) \quad (7)$$

where \bar{y}_3 and \bar{y}_1 are the sums of corrected data per genotype A_2A_2 and A_1A_1 , respectively, n_1 and n_3 are the number of animals with genotype A_1A_1 and A_2A_2 respectively.

Allele frequency : Given the genotypes of the base animals, allele frequency in the base population has a beta distribution (Janss et al., 1995):

$$\text{Sample } p_1^{[t-1]} \text{ from } f(p_1 \sim p_1^{N_1}(1-p_1)^{N_2}) \quad (8)$$

where N_1 is the number of A_1 alleles and N_2 is the number of A_2 alleles in genotypes of base animals.

Ten replicates of Gibbs chains of 50,000 cycles were run, using a spacing of 50 cycles, obtaining 1,000 Gibbs samples per chain and 10,000 samples in total for each trait. A burn-in period of 1,000 cycles was used to allow the Gibbs chains to reach the equilibrium.

Gibbs samples of the following parameters were directly obtained in each Gibbs cycle, variance components σ_u^2 , σ_{pe}^2 and σ_e^2 , additive major gene effect a , and allele frequencies p_1 and p_2 . These parameters allowed the computation in each Gibbs cycle of the within major gene heritability h^2 , repeatability r , genotypic frequencies $f(A_1A_1)$, $f(A_1A_2)$, $f(A_2A_1)$ and $f(A_2A_2)$ and major gene variance σ_m^2 . For the post Gibbs analysis of the samples, an analysis of variance was used to check for equality of chains. This test also yielded information about the dependency of the samples kept (Janss et al., 1997) within chain (convergence of chain).

It is well known that with Bayesian method, there is no construction of LR test or significance testing similar to what is applied regularly in standard statistical methods

such as ML or regression methods. Therefore, we were unable to apply MAGGIC software under two different hypotheses and construct likelihood ratio test to test for a significant contribution of major gene.

However, the highest posterior density regions, HPD, according to Box and Tiao (1973), based on a non-parametric density estimate using the average shifted histogram, were determined for all model parameters. The HPD guarantees that the density of each point within this region is equal or above the density of each point outside it. The HPD allows, e.g., the following reasoning: If the region for a variance component or a frequency includes the boundary value of zero then this parameter is not of importance for this particular trait. In the present investigation $1-\alpha = 0.95$ was used to construct the HPD. We could derive HPD regions to test for the significance of the major gene as in other studies (e.g. Miyake et al., 1999; Ilahi and Kadarmideen, 2004; Hagger et al., 2004).

For the post Gibbs analysis of the samples, an analysis of variance was used to check for equality of chains. This test also yields information about the dependency of the samples kept (Janss et al., 1997) within chain (convergence of chain). The marginal posterior means were used as estimator of the parameters.

Detection of major gene for bone lesions in pigs

Osteochondrosis (OC) in pigs is an abnormal bone development represented by ossification on the growth plates and cartilages at bone joints (Kadarmideen et al., 2004). Data on OC were obtained from SUISAG, a stock company for services in pig production including herd book, field and station tests and artificial insemination in Switzerland. Animals subject to performance test are sampled and recorded for osteochondrosis (OC) lesions during slaughter. Trained personnel in SUISAG conducted morphological examinations of front and hind leg bones of slaughtered pigs and manually recorded OC lesions with a score of: 1 = 'Normal' and 2 to 6 = 'mild to severely affected', depending on the lesion. There are 10 different parts of the bone scored OC lesions. To test the MIM methods proposed here, we applied Bayesian Linear Model (BALM) and Bayesian Threshold Models (BATM) to one of these lesions in front legs: condylus medialis humeri (CMH). A total of 1,291 animals were scored for OC lesions (score 1-5). The original data on CMH was a categorical score trait and were treated as normally distributed trait and analysed by BALM. To apply a BATM, the same data set was transformed to binary data sets as follows: animals with a score of '1' received a score of '0' (= 'healthy') and animals with a score of '2' and above received a score of 1 (= 'diseased'). This grouping of categories resulted in a binary data set (0 or 1) with 'healthy' and 'diseased' pigs. With this recoding, incidence

Table 2. True values of parameters and parameter estimates by maximum likelihood linear model segregation analyses for binary trait using two incidences on the observed scale: means and standard deviations of 100 replicates

| Parameters | True values ¹ (Incidence, 40%) | Incidence (15%) | | Incidence (40%) | |
|-----------------|--|---------------------------------|-----------------------------|---------------------------------|-----------------------------|
| | | Polygenic inheritance (H_0) | Mixed inheritance (H_1) | Polygenic inheritance (H_1) | Mixed inheritance (H_1) |
| μ | 0 | 0.190±0.022 | - | 0.410±0.072 | - |
| $\mu_{A_1A_1}$ | -0.945 | - | 0.023±0.004 | - | 0.053±0.023 |
| $\mu_{A_1A_2}$ | 0 | - | 0.050±0.005 | - | 0.096±0.018 |
| $\mu_{A_2A_2}$ | 0.945 | - | 0.852±0.011 | - | 0.882±0.019 |
| $f(A_1A_1)$ | 0.36 | - | 0.53±0.08 | - | 0.19±0.09 |
| $f(A_1A_2)$ | 0.48 | - | 0.38±0.10 | - | 0.51±0.14 |
| $f(A_2A_2)$ | 0.16 | - | 0.09±0.05 | - | 0.30±0.12 |
| σ_u^2 | 0.389 | 0.051±0.010 | 0.0005±0.000 | 0.091±0.021 | 0.001±0.00 |
| σ_{pe}^2 | 0.097 | 0.030±0.010 | 0.000 | 0.050±0.018 | 0.00 |
| σ_e^2 | 0.456 | 0.053±0.003 | 0.049±0.002 | 0.091±0.004 | 0.084±0.009 |
| σ_m^2 | 1.588 | - | 0.050±0.010 | - | 0.090±0.022 |
| Heritability | 0.11 | 0.38±0.07 | 0.010±0.004 | 0.39±0.08 | 0.012±0.007 |
| Repeatability | 0.14 | 0.60±0.02 | 0.010±0.004 | 0.60±0.02 | 0.012±0.007 |

¹ Obtained by transforming true values on NDJ scale to observed scale using Robertson and Lerner (1949) formula. True values for 15% incidence could not be derived.

of binary CMH lesion was 10.0%. Three replicates of Gibbs chains of 50,000 cycles were run, using a spacing of 50 cycles, obtaining 1,000 Gibbs samples per chain and 3,000 samples. A burn-in period of 1,000 cycles was used to allow the Gibbs chains to reach the equilibrium. Using these 3,000 samples, posterior mean (means) and standard deviations (SD) were computed.

RESULTS AND DISCUSSION

Liability data

The results of parameter estimates by segregation analyses using MLLM for liability data are given in Table 1. The empirical mean of the test statistic (likelihood ratio) comparing mixed and polygenic transmission models was about 165, greatly exceeding 13.3, the tabulated value of χ_4^2 distribution at 1% significance level. This has confirmed the true mixed genetic determinism of the simulated trait.

Estimated parameters under mixed inheritance (H_1) were similar to the true values of parameters used in the simulation except the major gene variance was underestimated. The estimated favorable allele A_2 frequency was 0.41. The major gene variance accounts for 73% of the total genetic variance of the trait and the difference between the values of the two homozygotes is 3.8 phenotypic standard deviation units of the trait. These estimates were very close to the true values of major gene parameters. However, the estimated parameters under polygenic inheritance (H_0) were overestimated, especially for the genetic and the permanent environmental variances. This is explained by the genetic model used in the simulation of data set: the major gene has a large additive effect on the trait. Moreover, under H_0 , the major gene effect was not taken into account to explain the genetic variability of the

analysed trait, which resulted in overestimation of genetic and permanent environmental variances. Similar findings were reported in the literature by Ilahi (1999) and Ilahi et al. (2000).

Categorical data

The results of parameters estimates by segregation analyses using MLLM for categorical data are given in Table 1. Under H_1 , genetic parameters were underestimated (e.g. true value of $h^2 = 0.41$ and $r = 0.52$ vs. estimates of 0.25 and 0.36, respectively). The estimated favorable allele A_2 frequency was 0.37. The major gene accounted for 74% of the total genetic variance of the trait and the difference between the values of the two homozygotes was 3.2 phenotypic standard deviation units.

It should be noted that, under both H_0 and H_1 hypotheses, the estimated variance components and genetic parameters obtained for discrete trait were lower than those for continuous trait (Table 1), as would be expected. These results are similar to those reported in the previous study (Le Roy and Elsen, 1991). Similar findings were also reported for QTL detection in non-normal traits (Rebai, 1997).

Binary data

Maximum likelihood linear model : Results of segregation analyses using this model for binary traits are given in Table 2. Using the formula proposed by Robertson and Lerner (1949), the true parameters values in the normal liability scale used in the simulation were transformed to the observed scale. Kadarmideen et al. (2000b) showed that this transformation also works well for transforming major gene/QTL effects on liability versus observed scales. Here, the transformation of the true values was applied to only

Table 3. True values of parameters and Bayesian marginal posterior means and marginal posterior standard deviations of parameters for binary trait¹ in a mixed inheritance model, based on 10,000 Gibbs samples from ten replicated chains

| Parameters | True values on liability scale | Incidence (15%) | Incidence (40%) |
|-----------------|--------------------------------|-----------------|-----------------|
| a | 3.50 | 4.305**±2.024 | 5.578**±2.651 |
| p_1 | 0.60 | 0.435±0.110 | 0.39**±0.196 |
| p_2 | 0.40 | 0.565±0.110 | 0.61**±0.196 |
| σ_a^2 | 1.44 | 1.711±0.618 | 1.231±0.528 |
| σ_{pe}^2 | 0.36 | 0.536±0.431 | 0.243±0.230 |
| σ_e^2 | 1.69 | 1.001±0.002 | 1.003±0.029 |
| σ_m^2 | 5.88 (1.59) | 9.110**±7.253 | 14.804**±12.206 |
| Heritability | 0.41 (0.11) | 0.52±0.117 | 0.49±0.093 |
| Repeatability | 0.52 (0.14) | 0.69±0.063 | 0.59±0.083 |

¹ Estimated parameters on NDL scale. ** Significant differences between replicated chains ($p < 0.01$).

high incidence 40% (Table 2), as the transformation using the low incidence 15% resulted in estimates outside the parameters space (e.g. $h^2 > 1$). For both incidences 15% and 40% under H_0 , the estimates of heritabilities and repeatabilities were the same and higher than their expected values on the observed scale. In the case of H_1 , however, the estimates are lower than the true values. This is due to the fact that under H_0 and for both incidences, the 3 variance components (σ_a^2 , σ_{pe}^2 and σ_e^2) estimates were lower than the true values and their scale of underestimation was generally the same, which is not the case under H_1 . Moreover, the MLLM did not allow the estimation of the permanent environmental variance (σ_{pe}^2). This may be due to the loss of variability and information when normal distributed data were truncated to 0/1 binary form (Xu and Atchley, 1996; Rebai, 1997; Kadarmideen et al., 2000b). In the study on segregation analyses for binary traits, Miyake et al. (2002) have also found similar problems in the estimation of variance components and to obtain a good convergence to true values. With low incidence (15%), there is an underestimation of the favorable genotype A_2A_2 frequency; however with high incidence (40%) there is an overestimation. This is similar to earlier findings that statistical power is lower and bias is higher for low incidence than those for intermediate incidence (Kadarmideen et al., 2000b).

Bayesian threshold model : The results of segregation analyses using Bayesian threshold model for binary traits are given in Table 3. These estimates of model parameters are based on 10,000 Gibbs samples from ten replicated chains. Tests for convergence of the Gibbs sampler were performed by comparison of multiple chain output using ANOVA on the total samples. These tests showed that Gibbs samples of parameters (for major gene effect, genotype frequencies and all variances) were not able to achieve a good stationary phase. The density estimates were higher than the true values. Similar finding on the overestimation of major gene variance was reported by Janss et al. (1995). Miyake et al. (2002) have also had problems to obtain a good convergence of the estimated

parameters for binary trait. Thaller et al. (1996) found that segregation analysis of binary traits by maximum likelihood method based on the infinitesimal mixed model is computationally feasible only when there are no relationships between parents and no additional random or fixed effects in the model.

Significance testing for major gene under BATM : The highest posterior density regions of major gene variances were from 2.25 to 26.50 and from 3.5 to 29.60 for 15 and 40% incidences respectively, which did not include the value zero. This confirmed the existence of a segregating major gene in both simulated binary datasets. The estimated polygenic and permanent environmental variances using 15% incidence were higher than those using 40% incidence. Conversely, the estimated major gene effect and variance were higher using 40% incidence than using 15% incidence (Table 3). This may be due to the fact that when the founder population has the same allele frequency at the major locus, this might infer a severe under or overestimation problems of parameters (Janss, 2003, personnel communication). Normally the magnitude of estimated parameters (of polygenic background and major gene) increase with the incidence (Kadarmideen et al., 2000b). In this study the genetic variability of simulated trait was almost explained by the major gene effect, this may have reduced the magnitude of parameter estimates of polygenic background when the incidence increases.

Comparison of liability vs. binary data : The expected value of test statistic (likelihood ratio) yielded by segregation analysis of normal data (165) would be expected to be higher than the likelihood ratio given by segregation analysis of the same normal data transformed to categorical or binary data. However, it is not the case in the present study. The empirical means of the test statistic were 511.40, 729.55 and 1,160.83 for categorical, 40 and 15% incidences binary data sets respectively. The assumptions of normality for discrete traits considerably increase the test statistic values and may therefore lead to false inference of a segregating major gene. This suggested that the MLLM is sensitive to deviation from normality. Similar finding was

Table 4. Estimated major gene and polygenetic parameters for osteochondral disease in pigs by mixed inheritance models, using Bayesian Linear Models (BALM) and Bayesian Threshold Models (BATM). Results are based on 10,000 Gibbs samples from three replicated chains

| Parameters | BALM to ordinal scores | | | | BATM to binary scores | | | |
|---------------------|------------------------|--------|------------------|------------------|-----------------------|--------|------------------|------------------|
| | Mean | SD | HPD _L | HPD _U | Mean | SD | HPD _L | HPD _U |
| a | 0.5870 | 0.0066 | 0.5744 | 0.6002 | 8.745 | 1.724 | 5.464 | 11.546 |
| p_1 | 0.2875 | 0.0299 | 0.2292 | 0.3464 | 0.396 | 0.032 | 0.335 | 0.459 |
| p_2 | 0.7125 | 0.0299 | 0.6541 | 0.7714 | 0.604 | 0.032 | 0.542 | 0.666 |
| σ_e^2 | 0.0129 | 0.0008 | 0.0113 | 0.0146 | 1.015 | 0.082 | 0.852 | 1.185 |
| σ_u^2 | 0.0009 | 0.0007 | 0.000 | 0.0023 | 1.301 | 0.438 | 0.541 | 2.038 |
| σ_m^2 | 0.0477 | 0.0130 | 0.0236 | 0.0734 | 37.815 | 13.137 | 12.747 | 62.253 |
| h^2 | 0.0653 | 0.0520 | 0.000 | 0.1611 | 0.556 | 0.098 | 0.347 | 0.700 |
| h^2 at major gene | 0.4016 | 0.0410 | 0.3200 | 0.5409 | 0.969 | 0.033 | 0.943 | 0.991 |

HPD_L: Highest probability density region at 95% level-lower limit.

HPD_U: Highest probability density region at 95% level-upper limit.

reported by Elsen and Le Roy (1990).

For liability data, the estimated polygenic heritabilities and repeatabilities were lowered from H_0 to H_1 , from 0.54 to 0.38 and from 0.80 to 0.51, respectively. Similar trends were observed for categorical trait (from 0.39 to 0.25 and from 0.74 to 0.36, respectively). This was expected due to the taking into account of major gene effect in H_1 (Ilahi et al., 2000). For binary traits, using MLLM, these estimates for both incidences were dramatically lowered from H_0 to H_1 . It decreased from 0.38 to 0.01 and from 0.60 to 0.01 for 15% incidence, and from 0.39 to 0.012 and 0.60 to 0.012 for 40% incidence. Moreover, even with the very large effect of the postulated major gene, the MLLM was not able to obtain accurate estimates for the major gene parameters. However, when using Bayesian threshold method (BATM) we obtained more reasonable and accurate estimates than those obtained with MLLM. This was in agreement with earlier studies (Yi and Xu, 2000). With MLLM method, estimated residual variance (σ_e^2) did not change from H_0 to H_1 for data on any distribution, there was an underestimation of polygenic variance (σ_u^2) for discrete traits, especially for binary traits with low incidence and non-estimability of the permanent environmental variance (σ_{pe}^2) for binary traits (both incidences). BATM estimates of polygenic effects under 40% incidence were more accurate than those estimates with 15% incidence. Since the loss of information due to the truncation of ND data to binary data is far larger than the loss of information when ND data is truncated to categorical data (5 categories), the bias in estimated parameters using MLLM is higher for binary scale than those for categorical scale. Similar results were reported for QTL detection (Rebai, 1997).

Application to bone diseases in pigs

Bayesian methods based on linear models (BALM) applied to CMF observed on original scales (scores 1-5) versus those based on threshold models (BATM) applied to transformed binary scales (0/1) are given in Table 4. Both methods showed a presence of major gene with significant

additive effect at the major gene (0.587 for BALM and 4.358 for BATM) and the very high additive genetic variance (0.0477 for BALM and 17.993 for BATM) compared to the polygenic variance for disease. Therefore the heritability at the major gene, h_m^2 , was much higher than the heritability at the polygenes, h_p^2 . The magnitude of the estimated parameters under BATM were higher than the same parameter estimated under BALM, as expected from the theory (on the continuous liability scale, variables have larger range than those on truncated binary data with probability range from 0 to 1.0). This has also been shown in the other studies (e.g. Kadarmideen et al., 2001, 2004). The frequency of allele that increases the incidence of the disease (A_2) was high 0.71 for BATM whereas it was 0.38 for BALM. In general, the probability of HPD region (lowest interval with zero) at 95% level was small for both methods. The posterior SDs for all parameters and for both methods was high, confirming that estimation of genetic parameters for categorical (ordinal or binary) traits are difficult to have good precision.

In general, both the methods developed here and application to bone disease data in pigs shows that segregation analysis could be used to first look for evidence of major gene affecting a trait, before embarking on establishment of resource populations for such a trait and genotyping animals for hundreds of DNA markers, as for example, shown in Kim et al. (2003) and Yan et al. (2004).

CONCLUSION

Maximum Likelihood Linear Model (MLLM) segregation analysis accurately estimated major gene parameters for ordinal or score data (e.g. disease severity), similar to those for normal data, under various simulation scenarios. The MLLM method applied to binary traits (e.g. healthy or diseased), however, failed to estimate all the model parameters under both polygenic and mixed inheritance models, whereas, Bayesian Threshold Models (BATM) detected segregating major gene and estimated its

parameters. BATM has more accuracy and less bias than MLLM for detecting major genes in binary traits. The application of developed Bayesian method under linear and threshold models to real bone diseases in pigs showed a presence of major gene with significant additive effect at the major gene and a very high additive genetic variance compared to the polygenic variance, indicating that this disease could be under the control of a major gene. The power and precision to detect major genes are generally lower for diseases scored on binary scale than those diseases that are more variable (ordinal or more than 2 categories). When incidence of diseases (on binary scale), is close intermediate levels (near 50%), segregation analysis had more accuracy and less bias compared to diseases with extreme or rare incidences. Normally distributed data (liability) usually had higher accuracy and lesser problems in inference/estimation of polygene and major gene parameters than categorical data.

ACKNOWLEDGEMENTS

Dr. J. M. Elsen and Dr. P. Le Roy are acknowledged for the provision of maximum likelihood segregation analysis software. Authors thank Dr. L. L. G. Janss for providing the MAGGIC software for Bayesian threshold segregation analysis and help to use this software. Authors thank SUISAG for providing data of pigs.

REFERENCES

- Bodin, L., M. San Cristobal-Gaudy, F. Lecerf, P. Mulsant, B. Bibé, D. Lajous, J. P. Belloc, F. Eycheime, Y. Amigues and J. M. Elsen. 2002. Segregation of major gene influencing ovulation in progeny of Lacaune meat sheep. *Genet. Sel. Evol.* 34:447-464.
- Box, G. E. P. and G. C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Reading-Mass. Addison-Wesley.
- Elston, R. C. and J. Stewart. 1971. A general model for the genetic analysis of pedigree data. *Hum. Hered.* 21:523-542.
- Falconer, D. S. and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. 4th edn, Longman, Harlow, London.
- Elsen, J. M. and P. Le Roy. 1990. Detection of major genes and determination of genotypes application to discrete variables. In *Proc. 4th World Congress. Genet. Appl. Livest. Prod.* Edinburgh, 23-27 July, 15:37-49.
- Gelfand, A. E. and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85:398-409.
- Geman, S. and D. Geman. 1984. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6:721-741.
- Gianola, D. 1982. Theory and analysis of threshold characters. *J. Anim. Sci.* 56:1079-1096.
- Gianola, D. and J. L. Folley. 1983. Sire evaluation for ordered categorical trait with a threshold models. *Genet. Sel. Evol.* 17:359-368.
- Guo, S., W. and E. A. Thompson. 1994. Monte Carlo estimation of mixed models for large complex pedigrees. *Biometrics* 50:417-432.
- Hagger, C., L. L. G. Janss, H. N. Kadarmideen and G. Stranzinger. 2004. Bayesian Inference on Major Loci in Related Multi-Generation Selection Lines of Laying Hens. *Poult. Sci.* 83:1932-1939.
- Hanset, R. and C. Michaux. 1985. On the genetic determinism of muscular hypertrophy in the Belgian White and Blue cattle breed: I. Experimental data. *Genet. Sel. Evol.* 15:201-224.
- Hill, W. G. and S. A. Knott. 1990. Identification of genes with large effects. In: *Advances in Statistical Methods for Genetic Improvement of Livestock*. Gianola, D., Hammond, K. Springer Verlag, New York. p. 477.
- Ilahi, H. 1999. Variabilité génétique du débit de traite chez les caprins laitiers. Ph.D Thesis, INRA-ENSA de Rennes.
- Ilahi, H., E. Manfredi, P. Chastin, F. Monod, J. M. Elsen and P. Le Roy. 2000. Genetic variability in milking speed of dairy goats. *Genet. Res.* 75:315-319.
- Ilahi, H. and H. N. Kadarmideen. 2004. Bayesian Segregation Analysis of Milk Flow in Swiss Dairy Cattle using Gibbs Sampling. *Genet. Sel. Evol.* 36:563-576.
- Janss, L. L. G., R. Thompson and J. A. M. Van Arendonk. 1995. Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theor. Appl. Genet.* 91:1137-1147.
- Janss, L. L. G., J. A. M. Van Arendonk and E. W. Brascamp. 1997. Bayesian statistical analyses for presence of single major genes affecting meat quality traits in crossed pig population. *Genetics* 145:395-408.
- Janss, L. L. G. 2004. MaGGic 4.1.: A package of subroutines for genetic analyses with Gibbs sampling.
- Kadarmideen, H. N., R. Thompson and G. Simm. 2000a. Linear and Threshold Model Genetic Parameters for Disease, Fertility and Milk Production in dairy cattle. *Anim. Sci.* 71:411-419.
- Kadarmideen, H. N., L. L. G. Janss and J. C. M. Dekkers. 2000b. Power of quantitative trait locus mapping for polygenic binary traits using generalized and regression interval mapping. *Genet. Res.* 76:305-317.
- Kadarmideen, H. N., R. Rekaya, D. Gianola. 2001. Genetic parameters for clinical mastitis in Holstein-Friesians: a Bayesian analysis. *Anim. Sci.* 73:229-240.
- Kadarmideen, H. N. and J. C. M. Dekkers. 2001. Generalized marker regression and interval QTL mapping methods for binary traits in half-sib family designs. *J. Anim. Breed. Genet.* 118:297-309.
- Kadarmideen, H. N. and L. L. G. Janss. 2003. Liability interval mapping of quantitative trait loci for complex polygenic binary diseases under gene by environmental interactions. *Proceedings of the XIX International Congress of Genetics*, Melbourne, Australia. Section: Animal breeding and Cloning. 6.C.0967: p. 134.
- Kadarmideen, H. N., D. Schwörer, H. Ilahi, M. Malek and A. Hofer. 2004. Genetics of osteochondral disease and its relationship with meat quality and quantity, growth and feed conversion traits in pigs. *J. Anim. Sci.* 82:3118-3127.
- Kim, J. W., S. I. Park and J. S. Yeo. 2003. Linkage Mapping and QTL on Chromosome 6 in Hanwoo (Korean cattle). *Asian-Aust. J. Anim. Sci.* 16:1402-1405.
- Knott, S. A., C. S. Haley and R. Thompson. 1991. Methods of

- segregation analysis for animal breeding data. a comparison of power. *Heredity* 68:299-311.
- Lee, D. H. 2002. Estimation of Genetic Parameters for Calving Ease by Heifers and Cows Using Multi-trait Threshold Animal Models with Bayesian Approach. *Asian-Aust. J. Anim. Sci.* 15:1085-1090.
- Le Roy, P., J. M. Elsen and S. Knott. 1989. Comparison of four statistical methods for detection of a major gene in a progeny test design. *Genet. Sel. Evol.* 21:341-357.
- Le Roy, P., J. Naveau, J. M. Elsen and P. Sellier. 1990. Evidence for a new major gene influencing meat quality in pigs. *Genet. Res.* 55:33-40.
- Le Roy, P. and J. M. Elsen. 1991. First statistical approaches of major gene detection with special reference to discrete traits. In: (Ed. J. M. Elsen, L. Bodin and J. Thimonier), 2nd International Workshop on Major Genes for Reproduction in Sheep. Toulouse, France, INRA Ed. Paris. 47:431-440.
- Miayke, T., T. Dogo, K. Moriya and Y. Sasaki. 1999. Bayesian analysis for existence of segregation of major genes affecting carcass traits in Japanese Black cattle population. *J. Anim. Breed. Genet.* 116:207-215.
- Miayke, T., Y. Sasaki, G. Dolf and C. Gaillard. 2002. Application of constraints on parameters in segregation analysis for binary traits using Gibbs sampling. In: (Ed. I. Hoeschele). In Proc. 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France. 32:609-612.
- Numerical Algorithms Group. 1990. The NAG Fortran Library Manual, NAG Ltd. Oxford.
- Piper, L. R. and B. M. Bindon. 1982. The Booroola Merino and the performance of medium non-peppin crosses at Armidale. In: *The Booroola Merino*, (Ed. L. R. Piper and B. M. Bindon), CSIRO, Melbourne, 9-20.
- Rebai, A. 1997. Comparison of methods for regression interval mapping in QTL analysis with non-normal traits. *Genet. Res.* 69:69-74.
- Ricordeau, G., J. Bouillon, P. Le Roy and J. M. Elsen. 1990. Déterminisme génétique du débit de traite au cours de la traite des chèvres. *INRA. Prod. Anim.* 3:121-126.
- Robertson, A. L. and I. M. Lerner. 1949. The heritability of all-or-none traits: viability of poultry. *Genetics* 34:395-411.
- Thaller, G., L. Dempfle and I. Hoeschele. 1996. Maximum likelihood analysis of rare binary traits under different modes of inheritance. *Genetics* 143:1819-1829.
- Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19:506-536.
- Xu, S. and W. R. Atchley. 1996. Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics* 143:1417-1424.
- Yan, X. M., J. Ren, H. S. Ai, N. S. Ding, J. Gao, Y. M. Guo, C. Y. Chen, J. W. Ma, Q. L. Shu and L. S. Huang. 2004. Genetic Variations Analysis and Characterization of the Fifth Intron of Porcine NRAMP1 Gene. *Asian-Aust. J. Anim. Sci.* 17:1183-1187.
- Yi, N. and S. Xu. 2000. Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics* 155:1391-1403.