# Bayesian genetic analysis of milk yield in Tunisian Holstein dairy cattle population

## Houcine Ilahi[1*], Mohamed Ben Hammouda[2], Mohamed Houcine Othmane[3]

[1]Université de Carthage, Institut Préparatoire aux Etudes d'Ingénieurs de Bizerte, Zarzouna, Tunisia
[2]Institution de Recherche et de l'Enseignement Supérieur Agricoles (IRESA), Ministère de l'Agriculture et de l'Environnement, Tunis, Tunisia
[3]Centre Régional des Recherches Agricoles de Sidi Bouzid, Sidi Bouzid, Tunisia
Email: *houcine.ilahi@ipeib.rnu.tn

## ABSTRACT

**The genetic determinism of 305-d milk yield and its genetic parameters were investigated in Tunisian Holstein dairy cattle population through Bayesian segregation analyses using a Monte Carlo Markov Chains (MCMC) method. Data included 49,709 records of 305-d milk yield collected between 1996 and 2003 from 114 dairy herds. The postulated major gene was assumed to be additive biallelic locus with Mendelian transmission probabilities and priors used for variance components were uniform. Gibbs sampling was used to generate a chain of 500,000 samples, which were used to obtain posterior means of genetic parameters. Estimated marginal posterior means ± posterior standard deviations of variance components of milk yield were 402866.28 ± 23629.97, 271256.66 ± 34477.83, 68276.83 ± 233027.62 and 1098855.75 ± 10009.52 for polygenic variance ($\sigma_u^2$), permanent environmental variance ($\sigma_{pe}^2$), major gene variance ($\sigma_G^2$) and error variance ($\sigma_e^2$), respectively. The main finding of this paper showed the postulated major locus was not significant, since the 95% highest posterior density regions (HPDs$_{95\%}$) of most major gene parameters included 0, and particularly for the major gene variance. Estimated transmission probabilities for the 95% highest posterior density regions (HPDs$_{95\%}$) were overlapped. Genetic parameters of 305-d milk yield were very similar under both mixed inheritance and polygenic models. These results indicated that the genetic determinism of milk yield in Tunisian Holstein dairy cattle population is purely polygenic. Based on 50,000 Gibbs samples, heritability and repeatability estimates using polygenic model were $h^2 = 0.22 \pm 0.012$ and $r = 0.38 \pm 0.006$, respectively.**

**Keywords:** Bayesian Analysis; Segregation; Major Gene; Genetic Parameters; Milk Yield; Holstein Dairy Cattle

## 1. INTRODUCTION

Milk yield is the most important source of income for dairy farmers where high producing cows are usually profitable [1]. The size of Holstein cow population has substantially increased over the recent years in Tunisia through the import of pregnant heifers and semen from temperate countries. In 2000, Holstein population accounted for more than 40% of the total cows in Tunisia [2]. Cows enrolled in the A4 official milk recording system since the 1960s, were about 10% of the total Holstein population in 2000 [3]. Alternate and owner farm recording systems are being highly encouraged in order to increase the number of cows enrolled in the national milk recording system. The data generated by the milk recording system is not sufficiently and not adequately used and valorized as well, especially because of a lack of genetic evaluation [2].

In livestock animal populations, large sizes of phenotypic observations (or records) are often available at low costs and it is worthwhile to use them to look for statistical evidence of major genes or quantitative trait loci (QTL) by statistical analysis. Segregation analysis is the most powerful statistical method to identify a single gene when DNA marker information is unavailable. With segregation analysis, it is possible to determine, using only phenotypic data sets, whether the inheritance of a given trait is controlled, at least in part, by a single gene with a large effect. Therefore, the existence of major genes has been investigated in several different studies in livestock species: Janss *et al.* [4] for various traits of Dutch Meishan crossbreds; Ilahi [5] and Ilahi *et al.* [6] for milking speed in dairy goats; Pan *et al.* [2] for somatic cell scores in dairy cattle; Hagger *et al.* [7] for selection response in laying hens; Ilahi and Kadarmideen [8] for milk flow in dairy cattle; Ilahi and Othmane [9,10] for milk yield in dairy sheep.

In a typical segregation analysis using pedigreed animal populations is impossible by analytical approaches due to

the existence of many (inbreeding) loops and due to the family sizes, which do not allow summing and integrating out genotypes and polygenic effects from the likelyhood or posterior density. This problem has been simplified by the development of Gibbs sampling, a Monte Carlo Markov Chain (MCMC) methodology [11] and its applications to livestock populations by Ilahi and Kadarmideen [8], Janss *et al.* [4,12] and Sorensen *et al.* [13].

The purpose of this paper was 1) to investigate whether a segregating major gene affects milk yield trait, and 2) to estimate the genetic parameters of milk yield in Tunisian Holstein dairy cattle using Bayesian analysis approach.

## 2. MATERIALS AND METHODS

### 2.1. Data

Data were from the official milk recording data sets provided by the Tunisian Genetic Improvement Center belonging to the Livestock and Pasture Office, Tunis. Data were recorded between 1996 and 2003. After editing on the herd size (≥20 records), and unreasonable production level (<1000 kg/lactation), the available data used for this paper consisted of 49,709 records of 305-d milk yield (with mean ± standard deviation: 6089.8 kg ± 1939.9) from 26,329 cows in 114 dairy herds. All pedigree information available was included in the analyses (4 generations). Thus, the pedigree included 56,238 animals with 4,670 different sires used within the pedigree.

### 2.2. Genetic Models

#### 2.2.1. Mixed Inheritance Model

To investigate the presence of a major gene for milk yield in Tunisian Holstein dairy cattle population, the following mixed inheritance model was applied:

$$y = X\beta + Zu + Qpe + ZWm + e$$

where $y$ is the vector of milk yield collected in 305 days with 2 times milking per day, $\beta$ is a vector of non-genetic fixed effects including: herd, year-season, and lactation number, $u$ is a random vector of individual polygenic effects, $pe$ is a random vector of permanent environmental effects, $W$ is a design matrix that contains the genotype of each individual (*i.e.*, *AA*, *AB*, *BB*), $m$ is the vector of genotype means (*i.e.*, –*a*, 0, *a*), $e$ is a random vector of residual effects, and $X$, $Z$ and $Q$ are incidence matrices relating the observations to their respective effects. In the term modelling the single major gene, both $W$ and $m$ are unknown and have to be estimated from data by using segregation analysis.

The number of levels of all effects included in the model and the number of animals in the pedigree are summarized in **Table 1**.

The major gene was modelled as an additive autosomal biallelic (*A* and *B*) locus with Mendelian transmission

**Table 1.** Number of levels of all effects used in the analyses.

| Effects | Number of levels |
| --- | --- |
| Herd | 114 |
| Year-season | 29 |
| Number of lactation (parity) | 7 |
| Permanent environmental effects | 26,329 |
| Animal in the pedigree | 56,238 |

probabilities. Allele *A* is defined to decrease the phenotypic value and allele *B* is defined to increase the phenotypic value (or favourable allele). With these two alleles *A* and *B*, with frequencies $p$ and $q = 1 - p$ where $p$ is the estimate of *A* allele frequency in the founder population in which the Hardy-Weinberg equilibrium was assumed, three genotypes *AA*, *AB* or *BA* and *BB* can be encountered, with genotype means $m = (-a, 0, a)$, where $a$ is the additive major gene effect.

Distributional assumptions for polygenic effects were, $u \sim N (0, A\sigma_u^2)$, where $A$ is the numerator relationship matrix. The distribution of the permanent environmental effects were, $pe \sim N (0, I\sigma_{pe}^2)$. Residual effects were assumed to be distributed as $e \sim N (0, I\sigma_e^2)$. $\sigma_u^2$, $\sigma_{pe}^2$ and $\sigma_e^2$ are polygenic, permanent environmental and residual variances, respectively. The relationship matrix of the full pedigree $A$ was used in the analyses. The variance attributable to the major gene ($\sigma_G^2$) was calculated as: $\sigma_e^2 = 2p(1-p)a^2$.

Uniform prior distributions were assumed in the range $(-\infty, +\infty)$ for non-genetic effects and effects at the major locus, in the range $(0, +\infty)$ for variance components, and in the range [0, 1] for allele frequencies [4].

Gibbs sampling algorithm with blocked sampling of genotypes $W$ was used for inference in the mixed inheritance model and implemented using the "iBay" software package version 1.46 developed by [14].

A single run of the Monte Carlo Markov Chains (MCMC) consisted of 520,000 Gibbs samples, with the first 20,000 samples used for burn-in period to allow the Gibbs chains to reach equilibrium. Thereafter each $10^{th}$ sample was collected to obtain 50,000 Gibbs samples in total.

From the mixed general model, marginal posterior densities of the following parameters were directly estimated in each Gibbs cycle: variance components $\sigma_G^2$, $\sigma_u^2$, $\sigma_{pe}^2$, and $\sigma_e^2$, additive effect at the major gene $a$, allele frequency $p$, and the Mendelian transmission probabilities. Using variance components for polygenes and major genes, following Janss [14], the heritabilities and repeatabilities were calculated as:

For heritability and repeatability:

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{pe}^2 + \sigma_e^2} \; ; \quad r = \frac{\sigma_u^2 + \sigma_{pe}^2}{\sigma_u^2 + \sigma_{pe}^2 + \sigma_e^2}$$

      **OPEN ACCESS**

For total heritability and repeatability:

$$h_t^2 = \frac{\sigma_u^2 + \sigma_G^2}{\sigma_u^2 + \sigma_G^2 + \sigma_{pe}^2 + \sigma_e^2} \; ; \; r_t = \frac{\sigma_u^2 + \sigma_G^2 + \sigma_{pe}^2}{\sigma_u^2 + \sigma_G^2 + \sigma_{pe}^2 + \sigma_e^2}$$

### 2.2.2. Polygenic Model

The aim of fitting a polygenic model to analyse again this data was to obtain genetic parameter estimates of 305-d milk yield in Tunisian Holstein dairy cattle population, and to compare them with those obtained using the mixed inheritance model, to check the mode of inheritance of this trait.

Based on the same statistical model used in mixed inheritance analysis described above, the variance components under a polygenic model of milk yield were estimated by Bayesian analysis, using the "iBay" software package version 1.46 as well [14].

## 3. RESULTS AND DISCUSSION

Marginal posterior means and standard deviations of parameter estimates of 305-d milk yield in Tunisian Holstein dairy cattle population using Bayesian segregation analyses implemented by Gibbs sampling are presented in **Tables 2** and **3**. These estimates were based on 50,000 Gibbs samples. Posterior marginal distributions of all variance components of milk yield trait are shown in **Figure 1**.

According to Box and Tiao [15], the highest posterior density regions (HPD), based on a non-parametric density estimate using the averaged shifted histogram technique [16], were obtained for all model parameters. These highest regions were constructed to include the smallest possible region of each sampled parameter values. The highest posterior density regions at 95% (HPDs$_{95\%}$) of the additive gene effect (*a*) and the variance at the major locus ($\sigma_G^2$) included zero (**Table 2**). The allele frequencies in the studied population were $p = 0.77$ and $q = 1 - p = 0.23$. The estimated polygenic variance ($\sigma_u^2 = 402866.28$) was significantly higher than the major gene variance ($\sigma_G^2 = 68276.83$). Janss *et al.* [17] and Miyake *et al.* [18] also suggested the use of the magnitude of the major gene variances as an indicator for the existence of a segregating major gene. Following Elston [19], the evidence of a significant segregating major gene in quantitative traits requires three conditions: statistical significance of the major gene component in the model, statistical differences among the transmission probabilities and these transmission probabilities are significantly different from an environmental model.

To check the statistical significance of the major gene component in the model, Janss [14] proposed to check the 95% highest posterior density region (HPD$_{95\%}$) of the postulated major gene variance: if the 95% HPD does not include zero (the postulated major gene is statistically significant) or includes zero (not significant).

The Mendelian transmission (probabilities 1, 1/2, and 0) was tested by checking if the highest posterior density regions at 95% (HPDs$_{95\%}$) were overlapped or not. Mendelian transmission probabilities for the 3 genotypes were estimated (**Table 3**) as suggested by Elston and Stewart [20]. These probabilities were parameterised to indicate the Mendelian transmission of the favourable allele, with probabilities of *B* allele transmission of 1, 1/2, and 0 for genotypes *BB*, *BA*, and *AA*, respectively.

Results in **Table 3** showed that the three estimated posterior means of Mendelian transmission probabilities were not significantly different, and as well their highest posterior density regions at 95% (HPDs$_{95\%}$) for the three genotypes were overlapped. Furthermore, the density of marginal posterior distribution for the major gene variance (**Figure 1**) was unimodal marginal density with mode = 0, suggesting the absence of a major gene for the analysed trait [4,21].

**Table 2.** Estimated marginal posterior means and marginal posterior standard deviations for fitted parameters from mixed model and left and right 95% highest posterior density regions (HPDs$_{95\%}$) for 305-d milk yield in Tunisian Holstein dairy cattle population, based on 50,000 Gibbs samples.

| Genetic parameters | Posterior means | Posterior standard deviations | Left HPD$_{95\%}$ | Right HPD$_{95\%}$ |
|---|---|---|---|---|
| Polyenic variance $\sigma_u^2$ | 402866.28 | 23629.97 | 357548.59 | 448555.38 |
| Permanent variance $\sigma_{pe}^2$ | 271256.66 | 34477.83 | 169029.39 | 323583.87 |
| Error variance $\sigma_e^2$ | 1098855.75 | 10009.52 | 1078760.75 | 1118025.63 |
| Major gene variance $\sigma_G^2$ | 68276.83 | 233027.62 | 0.00 | 877042.37 |
| Additive gene effect *a* | 213.44 | 379.60 | 0.00 | 1546.39 |
| Frequency of allele *A p* | 0.77 | 0.20 | 0.27 | 1.00 |
| Heritability $h^2$ | 0.23 | 0.02 | | |
| Repeatability *r* | 0.38 | 0.01 | | |

**Table 3.** Estimated marginal posterior means, left and right 95% highest posterior density regions (HPDs$_{95\%}$) for transmission probabilities using mixed model for 305-d milk yield in Tunisian Holstein dairy cattle population, based on 50,000 Gibbs samples.

| Transmission probability[*] | Posterior means | Left HPD$_{95\%}$ | Right HPD$_{95\%}$ |
|---|---|---|---|
| Pr(B\|BB) | 0.37 | 0.05 | 0.89 |
| Pr(B\|BA) | 0.17 | 0.00 | 0.40 |
| Pr(B\|AA) | 0.08 | 0.00 | 0.20 |

[*]Transmission probabilities, presented as the probabilities to inherit a *B* allele from *BB*, *BA*, and *AA* genotypes (Elston and Stewart (1971)).
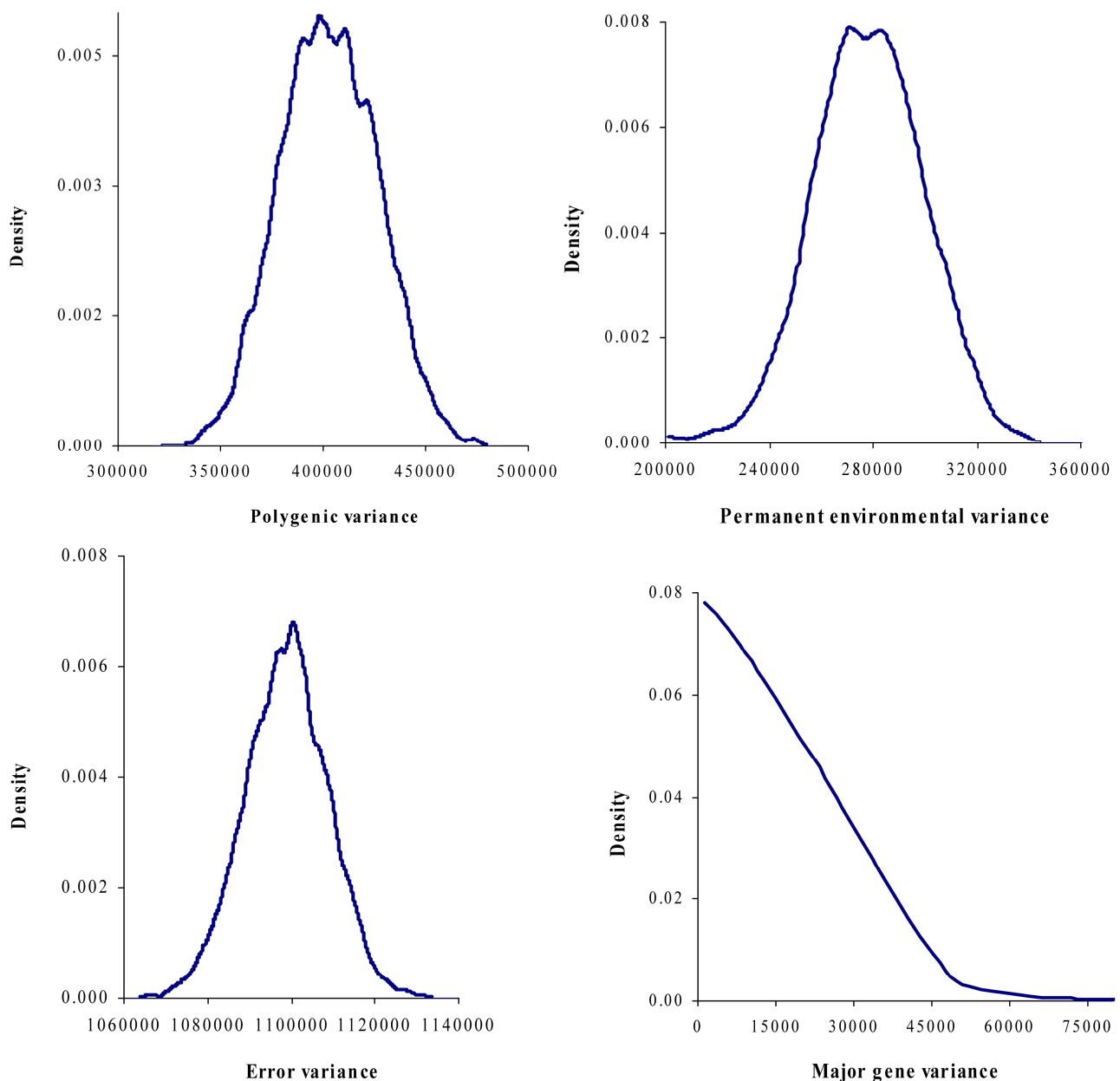
**Figure 1.** Marginal posterior distributions of polygenic variance, permanent environmental variance, error variance and major gene variance from mixed model of 305-d milk yield in Tunisian Holstein dairy cattle population.

According to these results obtained from segregation analysis via Gibbs sampling based only on phenotypic data sets, we can conclude that the postulated major gene was not significant and the genetic inheritance of milk yield in Holstein dairy cattle is polygenic.

Estimates of genetic parameters were consistent across models mixed and polygenic models (**Tables 2** and **4**). This finding confirmed again that the postulated major gene is not significant on milk yield in Holstein dairy cattle.

The variance components under polygenic model of 305-d milk yield were estimated by using Bayesian analy-

sis approach. These estimates were based on 50,000 Gibbs samples and were given in **Table 4**. Estimated heritability for 305-d milk yield using polygenic model (0.22 ± 0.012) was in accordance with previous estimates (0.25) on the same population obtained by Ben Gara *et al.* [22] and Hammami *et al.* [10]. However, this heritability value was still lower than those reported by several studies [23-25]. The low estimates for heritability might be explained by the limited production levels in Tunisian dairy cattle populations and the incomplete and/or inaccurate pedigree information on imported semen of some sires [3]. Repeatability estimated using polygenic model

**Table 4.** Estimated marginal posterior means and marginal posterior standard deviations for variance components from polygenic model and left and right 95% highest posterior density regions (HPDs$_{95\%}$) for 305-d milk yield in Tunisian Holstein dairy cattle population, based on 50,000 Gibbs samples.

| Genetic parameters | Posterior means | Posterior standard deviations | Left HPD$_{95\%}$ | Right HPD$_{95\%}$ |
|---|---|---|---|---|
| Polyenic variance $\sigma_u^2$ | 404241.44 | 22962.98 | 359199.19 | 449280.47 |
| Permanent variance $\sigma_{pe}^2$ | 282589.06 | 19751.26 | 243750.80 | 320823.69 |
| Error variance $\sigma_e^2$ | 1099467.38 | 9994.57 | 1080144.63 | 1119170.00 |
| Polygenic heritability $h^2$ | 0.22 | 0.012 | - | - |
| Repeatability $r$ | 0.38 | 0.006 | - | - |

$(0.38 \pm 0.006)$ was comparable to those found in the literature [22,26-28].

## 4. CONCLUSION

The mode of inheritance and genetic parameters of 305-d milk yield in Tunisian Holstein dairy cattle population were investigated and estimated under mixed inheritance and polygenic models via Gibbs sampling. Milk yield heritability was relatively low because of the low production levels in such a population, and for lack of information on pedigree of some imported sires. Our results based only on phenotypic data showed no existence of major gene and the mode of inheritance for milk yield in Tunisian Holstein dairy cattle population is then purely polygenic.

## 5. ACKNOWLEDGEMENTS

## REFERENCES

[1] Degroot, B.J., Keown, J.F., Van Vleck, L.D., *et al*. (2002) Genetic parameters and response of linear type, yield traits, and somatic cell scores to divergent selection for predicted transmitting ability for type in Holsteins. *Journal of Dairy Science*, **85**, 1314-1323. doi:10.3168/jds.S0022-0302(02)74227-6

[2] Hammami, H., Rekik, B., Soyeurt, H., *et al*. (2008) Genetic parameters for Tunisian Holsteins using a test-day random regression model. *Journal of Dairy Science*, **91**, 2118-2126. doi:10.3168/jds.2007-0382

[3] Rekik, B., Ben Gara, A., Ben Hammouda, M., *et al*. (2003) Fitting lactation curves of dairy cattle in different types of herds in Tunisian. *Livestock Production Science*, **83**, 309-315. doi:10.1016/S0301-6226(03)00028-9

[4] Janss, L.L.G., Thompson, R. and Van Arendonk, J.A.M. (1995) Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theoretical and Applied Genetics*, **91**, 1137-1147. doi:10.1007/BF00223932

[5] Ilahi, H. (1999) Variabilité génétique de débit de traite chez les caprins laitiers. Thèse de doctorat. ENSA, Rennes.

[6] Ilahi, H., Manfredi, E., Chastin, P., *et al*. (2000) Genetic variability in milking speed of dairy goats. *Genetical Research*, **75**, 315-319. doi:10.1017/S001667230000450X

[7] Hagger, C., Janss, L.L.G., Kadarmideen, H.N., *et al*. (2004) Bayesian inference on major loci in related multi-generation selection lines of laying hens. *Poultry Science*, **83**, 1932-1939.

[8] Ilahi, H. and Kadarmideen, H.N. (2004) Bayesian segregation analysis of milk flow in Swiss dairy cattle using Gibbs sampling. *Genetics Selection Evolution*, **36**, 563-576. doi:10.1186/1297-9686-36-5-563

[9] Ilahi, H. and Othmane, M.H. (2011) Complex segregation analysis of total milk yield in Churra dairy ewes. *Asian-Australian Journal of Animal Science*, **24**, 330-335.

[10] Ilahi H. and Othmane, M.H. (2011) Bayesian segregation analysis of test-day milk yield in Tunisian Sicilo-Sarde dairy sheep. *Journal of Animal and Feed Science*, **20**, 161-170.

[11] Guo, S.W. and Thompson, E.A. (1992) Monte Carlo method for combined segregation and linkage analysis. *Journal of Human Genetics*, **51**, 1111-1126.

[12] Janss, L.L.G., Van Arendonk, J.A.M. and Brascamp, E.W. (1997) Bayesian statistical analyses for presence of single major genes affecting meat quality traits in crossed pig population. *Genetics*, **145**, 395-408.

[13] Sorensen, D., Anderson, S., Jensen, J., *et al*. (1994) Inferences about genetic parameters using the Gibbs sampler, *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Guelph, 7-12 August 1994, 321-328.

[14] Janss, L.L.G. (2008) iBay manual version 1.46. Janss Bioinformatics, Lieden.

[15] Box, G.E.P. and Tiao, G. (1973) Bayesian inference in statistical analysis. Reading Addison-Wesley, New York.

[16] Scott, D.W. (1992) Multivariate density estimation. Wiley and Sons, New York. doi:10.1002/9780470316849

[17] Janss, L.L.G. (1998) "MAGGIC" a package of subroutines for genetic analyses with Gibbs sampling. *Proceeding of 6th World Congress on Genetics Applied to Livestock Production*, Armidale, 6-11 January 1998, University of New England, 459-460.

[18] Miyake, T., Gaillard, C., Moriya, K., *et al*. (1999) Accuracy of detection of major genes segregating in outbred population by Gibbs sampling using phenotypic values of quantitative traits. *Journal of Animal Breeding and Genetics*, **116**, 281-288. doi:10.1046/j.1439-0388.1999.00197.x

[19] Elston, R.C. (1980) Segregation analysis. Current developments in anthropological genetics. Mielke, J.H. and Crawford, M.H., Eds., Publishing Corporation, 327-354.

doi:10.1007/978-1-4613-3084-4_11

[20] Elston, R.C. and Stewart, J.M. (1971) A general model for the genetic analysis of pedigree data. *Human Heredity*, **21**, 523-542. doi:10.1159/000152448

[21] Pan, Y., Boettcher, J. and Gibson, J. (2001) Bayesian segregation analyses of somatic cell scores of Ontario Holstein cattle. *Journal of Dairy Science*, **84**, 2796-2802. doi:10.3168/jds.S0022-0302(01)74735-2

[22] Ben Gara, A., Rekik, B. and Bouallègue, M. (2006) Genetic parameters and evaluation of Tunisian dairy cattle population for milk yield by Bayesian and BLUP analyses. *Livestock Production Science*, **100**, 142-149. doi:10.1016/j.livprodsci.2005.08.012

[23] Haile-Mariam, R., Bowman, J.P. and Goddar, M.E. (2003) Genetic and environmental relationship among calving interval, survival, persistency of milk yield and somatic cell count in dairy cattle. *Livestock Production Science*, **116**, 189-200. doi:10.1016/S0301-6226(02)00188-4

[24] Jiang, X.P., Lieu, G.Q., Wang, C., *et al*. (2004) Milk trait heritability and correlation with heterozygosity in yak. *Journal of Applied Genetics*, **45**, 215-224.

[25] Weigel, K.A., Rekaya, R., Zwald, N.R., *et al*. (2001) International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records. *Journal of Dairy Science*, **84**, 2789-2795. doi:10.3168/jds.S0022-0302(01)74734-0

[26] Bakir, G., Kaygisiz, A. and Ulher, H. (2004) Estimates of genetic parameters of milk yield in brown Swiss and Holstein Friesian cattle. *Pakistan Journal of Biological Sciences*, **7**, 1198-1201. doi:10.3923/pjbs.2004.1198.1201

[27] Chauhan, V.P.S. and Hayes, J.F. (1991) Genetic parameters for first lactation milk production and composition for Holsteins using multivariate restricted maximum likelihood. *Journal of Dairy Science*, **74**, 603-615. doi:10.3168/jds.S0022-0302(91)78207-6

[28] Misztal, I., Lawlor, T.J., Short, T.H., *et al*. (1992) Multiple trait estimation of variance components of milk yield and type traits using an animal model. *Journal of Dairy Science*, **75**, 544-551. doi:10.3168/jds.S0022-0302(92)77791-1