Introduction

Genetic improvement of dairy cattle would increase benefits of dairy farmers. The size of Holstein cow population has substantially increased over recent years in Tunisia through the importation of cattle and semen from developed countries (USA, the Netherlands, and Germany). Hammami et al. (2007) and Ben Zaabza al. (2016) reported that 60% of all inseminations of dairy cows in Tunisia used Holstein semen. Recently, more attention has been given and placed on milk quality traits in breeding programmes. Estimates of genetic parameters for milk yield in dairy cows are abundant in the literature (Hammami et al., 2008a, 2009a, Ben Gara et al., 2006, 2012; Ilahi et al., 2012). However, investigation of genetic parameters of milk quality traits is lacking. Cows enrolled in the A4 official milk recording system since the 1960s, were about 10% of the total Holstein population in 2000 (Rekik et al., 2003). Alternate and owner farm recording systems are being highly encouraged to increase the number of Holstein 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 lack of genetic evaluation (Hammami et al., 2008b).

Abstract: Data including 15 343 records of 5 881 first parity Holstein cows collected from 2005 to 2017 in 118 herds located in north of Tunisia were analysed for fat and protein contents. Bayesian segregation analyses using a Monte Carlo Markov Chains (MCMC) method were used to investigate genetic determinism of both traits and to estimate its genetic parameters. The postulated major gene in both traits 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 600 000 samples, which were used to obtain posterior means of genetic parameters. Estimated marginal posterior means ± posterior standard deviations of variance components of fat and protein contents were 1.63 ± 0.37 0.60 ± 0.57 0.29 ± 0.32 0.03 ± 0.01 0.77 ± 0.83 0.33 ± 0.25 and 9.30 ± 0.20 3.32 ± 0.05 for polygenic variance (σ²), permanent environmental variance (σ²e), major gene variance (σ²g) and error variance (σ²e), respectively. Results showed the postulated major locus for both traits fat and protein contents was not significant, since the 95% highest posterior density regions (HPDs) of most major gene parameters included 0, and particularly for the major gene variances. Estimated transmission probabilities for the 95% highest posterior density regions (HPDs) of both traits were overlapped. Genetic parameters estimates of fat and protein contents were very similar under both mixed inheritance and polygenic models. These results indicated that the genetic determinism of fat and protein contents in Tunisian Holstein dairy cows is purely polygenic. Based on 30 000 Gibbs samples, polygenic heritability estimates for fat and protein contents were 0.18 and 0.20, respectively. The corresponding repeatability estimates were 0.24 and 0.22, respectively.

Keywords: Holstein cows, fat content, protein content, Bayesian segregation analysis, major gene; genetic parameters.
Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling

In farm animal populations, large sizes of phenotypic observations/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 certain 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. (1995) for various traits of Dutch Meishan crossbreds; Ilahi (1999) and Ilahi et al. (2000) for milking speed in dairy goats; Pan et al. (2001) for somatic cell scores in dairy cattle; Hagger et al. (2004) for selection response in laying hens; Ilahi and Kadarmideen (2004) for milk flow in dairy cattle; Ilahi and Othmane (2011a, b) for milk yield in dairy sheep; Ilahi et al. (2012) for milk yield in dairy cattle.

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 to sum and integrate out genotypes and polygenic effects from the likelihood or posterior density. This problem has been simplified by the development of Gibbs sampling, a Monte Carlo Markov chain (MCMC) methodology (Guo and Thompson, 1992) and its applications to livestock populations by Sorensen et al. (1994), Janss et al. (1995), Janss et al. (1997), and Ilahi and Kadarmideen (2004).

The aim of this paper was to investigate whether a segregating major gene affects fat and protein contents and to estimate the genetic parameters of fat and protein contents in Tunisian Holstein dairy cows via Bayesian analysis approach.

**MATERIALS AND METHODS**

**Data**

Data were provided by the Tunisian Genetic Improvement Center, Livestock and Pasture Office, Tunis. Original data from the official recording database were recorded between 2005 and 2017 in 118 dairy herds located in north of Tunisia. After editing on the herd size (≥ 5 records), and records only that included fat and protein contents were retained. Cows without pedigree information were discarded. Finally, data set included 15 343 records of 5 881 first parity from 5 042 cows for fat (35.38 g/kg ± 3.80) and protein (36.00 g/kg ± 3.98) contents were used for genetic analysis. Pedigree was traced as far back as possible and all pedigree information available was included in the analyses. Thus, the pedigree included 15 171 animals with 1 048 different sires.

**Genetic Models**

**Mixed Inheritance Model**

To estimate variance components and to investigate the presence of a major gene for fat and protein contents in Holstein dairy cows, the following mixed inheritance single trait model was used:

\[ y = X\beta + Zu + Qpe + ZWm + e \]

where \( y \) is the vector of observations (fat and protein contents), \( \beta \) is a vector of non-genetic fixed effects including: herd, year of calving, 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.
Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling

modelling the single 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 shown in Table 1.

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

<table>
<thead>
<tr>
<th>Effects</th>
<th>Number of levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herd</td>
<td>118</td>
</tr>
<tr>
<td>Year of calving</td>
<td>13</td>
</tr>
<tr>
<td>Lactation number</td>
<td>7</td>
</tr>
<tr>
<td>Permanent environmental</td>
<td>7112</td>
</tr>
<tr>
<td>Animal in the pedigree</td>
<td>15171</td>
</tr>
</tbody>
</table>

The major gene was modelled as an additive autosomal biallelic (A and B) locus with Mendelian transmission 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, $u ~ N \left(0, \sigma_u^2\right)$, where A is the numerator relationship matrix. The distribution of the permanent environmental effects were, $pe ~ N \left(0, \sigma_{pe}^2\right)$. Residual effects were assumed to be distributed as $e ~ N \left(0, \sigma_e^2\right)$. 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: $\sigma_G^2 = 2p(1-p)a^2$.

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 (Janss et al. 1995).

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 (Janss 2008).

A single run of the Monte Carlo Markov Chains (MCMC) consisted of 620 000 Gibbs samples, with the first 20 000 samples used for burn-in period to allow the Gibbs chains to reach equilibrium. Thereafter each 20$^{th}$ was collected to obtain 30 000 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 $\sigma_u^2, \sigma_{pe}^2$ 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 (2008), the heritabilities and repeatabilities were computed as:

For heritability and repeatability: $h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{pe}^2 + \sigma_e^2}$; $r = \frac{\sigma_G^2 + \sigma_{pe}^2}{\sigma_u^2 + \sigma_{pe}^2 + \sigma_e^2}$

For total heritability and repeatability: $h_t^2 = \frac{\sigma_u^2 + \sigma_G^2}{\sigma_u^2 + \sigma_G^2 + \sigma_{pe}^2 + \sigma_e^2}$; $r_t = \frac{\sigma_{pe}^2 + \sigma_G^2}{\sigma_u^2 + \sigma_G^2 + \sigma_{pe}^2 + \sigma_e^2}$
Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling

**Polygenic Model**

The objective of fitting a polygenic model to analyse again this data was to obtain genetic parameter estimates of fat and protein contents in Tunisian Holstein dairy cows, and to compare them with those obtained using the mixed inheritance model, to check the mode of inheritance (genetic determinism) of these both analysed traits.

Based on the same statistical model used in mixed inheritance analysis described above, the variance components under a polygenic model of fat and protein contents were estimated by Bayesian analysis, using the ‘iBay’ software package version 1.46 as well (Janss, 2008).

**RESULTS AND DISCUSSION**

Marginal posterior means and standard deviations of parameters estimates of fat and protein contents using Bayesian segregation analyses implemented by Gibbs sampling are shown in Tables 2 and 3. These estimates are based on 600 000 Gibbs samples. Posterior marginal distributions of all variance components of both traits are presented in Figures 1 and 2.

**Table 2.** Estimated marginal posterior means and marginal posterior standard deviations (SD) for fitted parameters from mixed model and left and right 95% highest posterior density regions (HPDs) for fat and protein content in Holstein dairy cows, based on 30 000 Gibbs samples.

| Genetic parameters | Fat content | | | | | Protein content | | | |
|-------------------|-------------|-----------------|-----------------|-------------|-----------------|-----------------|-----------------|-------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                   | Mean        | SD              | Left HPD 95%    | Right HPD 95%  | Mean         | SD              | Left HPD 95%    | Right HPD 95%  |
| Polyenic variance  | 1.63        | 0.37            | 1.02            | 3.48          | 0.60          | 0.57            | 0.49            | 0.70          |
| Permanent variance | 0.29        | 0.32            | 0.01            | 1.013         | 0.03          | 0.01            | 0.01            | 0.06          |
| Error variance     | 9.30        | 0.20            | 8.92            | 9.72          | 3.32          | 0.05            | 3.22            | 3.43          |
| Major gene variance| 0.77        | 0.83            | 0.00            | 2.51          | 0.33          | 0.25            | 0.00            | 0.80          |
| Additive gene effect| 4.33        | 2.08            | 0.00            | 6.21          | 4.18          | 0.38            | 0.01            | 4.93          |
| Frequency of allele Ap| 0.89        | 0.19            | 0.40            | 1.00          | 0.99          | 0.01            | 0.97            | 1.00          |
| Heritability h²    | 0.20        | 0.02            | -               | 0.22          | 0.04          | -               | 1               |
| Repeatability r    | 0.23        | 0.01            | -               | 0.23          | 0.06          | -               | 1               |

**Table 3.** Estimated marginal posterior means, left and right 95% highest posterior density regions (HPDs) for transmission probabilities using mixed model for fat and protein content in Holstein dairy cows, based on 30 000 Gibbs samples.

| Transmission probability* | Fat content | | | | | Protein content | | | |
|---------------------------|-------------|-----------------|-----------------|-------------|-----------------|-----------------|-----------------|-------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                           | Mean        | Left HPD 95%    | Right HPD 95%   | Mean        | Left HPD 95%    | Right HPD 95%   |
| Pr(B|BB)                     | 0.89        | 0.40            | 1.00            | 0.99        | 0.98            | 1.00            |
| Pr(B|BA)                     | 0.58        | 0.10            | 0.87            | 0.75        | 0.57            | 0.94            |
| Pr(B|AA)                     | 0.29        | 0.00            | 0.65            | 0.41        | 0.00            | 0.77            |

* Transmission probabilities, presented as the probabilities to inherit a B allele from BB, BA, and AA genotypes (Elston and Stewart (1971)).
Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling

Figure 1. Marginal posterior distributions of polygenic variance, permanent environmental variance, error variance and major gene variance from mixed model of fat content in Tunisian Holstein dairy cows.
Figure 2. Marginal posterior distributions of polygenic variance, permanent environmental variance, error variance and major gene variance from mixed model of protein content in Tunisian Holstein dairy cows.
Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling

According to Box and Tiao (1973), the highest posterior density regions (HPD), based on a non-parametric density estimate using the averaged shifted histogram technique (Scott, 1992), were obtained for all model parameters and for both trait. These highest regions were constructed to include the smallest possible region of each sampled parameter values. The highest posterior density regions at 95% (HPD_{95%}) of the additive gene effect (a) and the variance at the major locus (\sigma^2_m) of fat and protein contents included zero (Table 2). The estimated polygenic variances were 1.63 and 0.60 for fat and protein contents respectively, these estimates were significantly higher than the major gene variances 0.77 and 0.33 for fat and protein contents respectively. Janss et al. (1997) and Miyake et al. (1999) also suggested the use of the magnitude of the major gene variances as an indicator for the existence of segregating a major gene. However, following Elston (1980), the evidence of a significant segregating major gene in quantitative trait 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 (1998) proposed to check the 95% highest posterior density region (HPD_{95%}) of the postulated major gene variance: if the 95% HPD not include zero (the postulated major gene is statistically significant) or include zero (not significant). The Mendelian transmission (probabilities 1, 1/2, and 0) for both traits was tested by checking if the highest posterior density regions at 95% (HPD_{95%}) were overlapped or not. Mendelian transmission probabilities for the 3 genotypes were estimated (Table 3) as suggested by Elston and Stewart (1971). These probabilities are 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.

Table 3, shows the estimates posterior means of Mendelian transmission probabilities were not significantly different, and as well their highest posterior density regions at 95% (HPD_{95%}) for the three genotypes were overlapped. Furthermore, the density of marginal posterior distribution for the major gene variances as shown Figure 1 and 2 were unimodal marginal density with mode = 0, suggested the absence of a major gene for both analysed traits (Janss et al., 1995; Pan et al., 2001).

Following 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 both traits fat and protein contents in Holstein dairy cows is polygenic.

Tables 2 and 4 shows the estimates of genetic parameters are consistent across models mixed and polygenic models. This finding confirmed again that the postulated major gene is not significant on fat and protein contents in Holstein dairy cows.

**Table 4.** Estimated marginal posterior means and marginal posterior standard deviations (SD) for variance components from polygenic model and left and right 95% highest posterior density regions (HPD_{95%}) for fat and protein content in Holstein dairy cows, based on 30 000 Gibbs samples.

<table>
<thead>
<tr>
<th>Genetic parameters</th>
<th>Fat content</th>
<th>Protein content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Polyenic variance ( \sigma^2_p )</td>
<td>2.23</td>
<td>0.30</td>
</tr>
<tr>
<td>Permanent variance ( \sigma^2_p )</td>
<td>0.84</td>
<td>0.28</td>
</tr>
<tr>
<td>Error variance ( \sigma^2_e )</td>
<td>9.51</td>
<td>0.18</td>
</tr>
<tr>
<td>Polygenic heritability ( h^2 )</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>Repeatability ( r )</td>
<td>0.24</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling

The variance components under polygenic model of fat and protein contents were estimated by using Bayesian analysis approach. These estimates were based on 30 000 Gibbs samples and were shown in Table 4.

Heritability estimates for fat and protein contents using polygenic model were $0.18 \pm 0.02$ and $0.20 \pm 0.01$, respectively, which were in agreement with estimates obtained by Abdullahpour et al. (2010) and Abdullahpour et al. (2010). However, the results of this study were lower than those reported in the literature (Rzewuska and Strabel, 2013; Missanjo et al., 2013; Sneddon et al., 2015). The repeatability estimates were $0.24 \pm 0.01$ and $0.22\pm 0.01$ for fat and protein contents, respectively. These estimates were lower than those reported by Boujenane (2002); Missanjo et al., (2013) and Sneddon et al., (2015). The low estimates of genetic parameters might be explained by limited production levels in Tunisian dairy cattle population and incomplete and/or inaccurate pedigree information on imported semen of some sires (Rekik et al. 2003).

CONCLUSION

The mode of inheritance and genetic parameters of fat and protein contents in Tunisian Holstein dairy cows were investigated and estimated under mixed inheritance and polygenic single trait models via Gibbs sampling. The finding of this paper showed no existence of major gene and the genetic determinism of fat and protein contents is purely polygenic.

Estimates of genetic parameters of both traits were generally lower than those commonly reported in the literature, which may be due to limited production level of Tunisian dairy cattle populations, and the incomplete pedigree information on the fitted data.

Acknowledgments

Authors thank Dr. L.L.G. Janss for supplying iBay software for the analysis. Authors also thank Tunisian Livestock and Pasture Office for providing the data.

REFERENCES

Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling


Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling


Citation: Houcine Ilahi, Mohamed Ben Yedder, Mouna Hedhly, “Estimation of Genetic Parameters of Fat and Protein Contents in Tunisian Holstein Dairy Cows Using Gibbs Sampling”. American Research Journal of Genetics; vol 1, no. 1: 1-10

Copyright © Houcine Ilahi, Mohamed Ben Yedder, Mouna Hedhly, This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.