Genetic and environment info in goat milk FTIR spectra

BINYAM DAGNACHEW¹ AND TORMOD ÅDNØY¹
Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences

Background

Milk is a complex mixture of several thousand kinds of molecules (Harding, 1995); however, only a few major milk components are included in the current animal breeding programmes. The need for costly and time consuming chemical analysis to determine them limited their current use. Consequently, the variability of a large part of milk composition remains unknown. In a recent study, the genetic variability of cows’ milk components based on mid-infrared spectra (MIR) are presented (Soyeurt et al., 2010). The study indicated that some regions of the MIR spectra are more heritable than others.

Fourier Transform Infrared (FTIR) has a wide range of application, among others quantification of milk components (Iñón et al., 2004). Genetic evaluation of Norwegian dairy goats is based on major milk components (% fat, protein, lactose, etc) in samples of individual goats predicted by FTIR spectra obtained by Milkoscan™ Combfoss 6500 instruments (Foss, Hillerød, Denmark). These components are predicted from different calibration equation applied to the spectral data. The milk spectra could reflect both genetic and environmental variations of several thousand other milk components. Until now the genetic variability of the goat milk FTIR spectra are studied indirectly as they contribute to the major milk components they have been predict. The objective of this study is to investigate genetic and environmental variability of goat milk FTIR spectra directly.

Materials and Methods

FTIR spectra data: Raw milk FTIR spectra were obtained from TINE¹ database. For this study two years (2007 and 2008) of raw spectral data were collected. It included 74,858 observations of 28,260 goats. Due to the water, both the O-H stretching region (approximately between 1600 cm⁻¹ and 1680 cm⁻¹) and the O-H bending regions (above 3025 cm⁻¹) are more or less opaque to infrared light (Iñón et al., 2004) in milk samples. These two regions (accounting for 536 spectral data points) were omitted and the remaining 524 spectral data points (926 cm⁻¹ – 1597 cm⁻¹ and 1682 cm⁻¹ – 3025 cm⁻¹) were selected for further analysis.

¹ TINE is Norway’s largest producer, distributor, and exporter of dairy products
**Definition of new traits**: direct genetic analysis of all data points simultaneously was not possible with the current available methods. To reduce the dimension of spectral variables, principal component analysis (PCA) was carried out on the selected 524 spectral data points. The goal of PCA is to find a set of few principal components (PCs) that explains as much of the variance of the original variables as possible (Martens and Næs, 1989). The PCs are formulated as a linear combination of the original variables. The contributions of the original variables to each PC are given by a set of eigenvectors (loadings) and the amount of variance retained by each PC is given by its eigenvalue.

Let $Y_{(N \times 524)}$ be a matrix of the selected spectral data where $N$ is the total number of observations. PCA can be shown as:

$$Y_{(N \times 524)} = T_{(N \times M)}E'_{(M \times 524)} + F_{(N \times 524)}$$  \[1\]

Where $M$ is the number of PCs extracted, $T_{(N \times M)}$ represents score matrix of the M PCs, $E'_{(M \times 524)}$ represents transpose of loading matrix of the M PCs and $F_{(N \times 524)}$ represents un-modelled residual after the M PCs extracted from $Y_{(N \times 524)}$. PCA was performed on the correlation matrix of $Y_{(N \times 524)}$ using PROC PRINCOMP of SAS (SAS, 2004). Scores, $T_{(N \times M)}$, are considered as new traits and from now onwards they are referred to as PC traits.

**Model**: The PC traits ($T_{(N \times M)}$) were precorrected for fixed effect of herd region (12 levels); fixed effect of age at kidding (5 levels); fixed effect of number of kids (3 levels); fixed effect of stage of lactation (4 co-variables); and fixed effect of kidding season (3 levels). A multi-trait testday model was applied to the precorrected PC traits:

$$t_c = Xb + Za + Qp + Wh + e$$  \[2\]

Where $t_c$ is the vector of precorrected PC traits; $b$ is the vector of fixed effects; $a$ is the vector of genetic effect; $p$ is the vector of permanent environment effect of animals; $h$ is the vector of random herd-test-day effect; $e$ is the vector of random residual; $X$, $Z$, $Q$ and $W$ are the corresponding design matrices.

**Variance component estimation**: Variance components estimation was done using the multivariate AI-REML algorithm of WOMBAT (Meyer, 2007). Estimated variance components for the PC traits were back transformed to original spectral variables.
Results and Discussion

PCA on 524 spectral data points resulted in 8 principal components (PCs), which explained 98.98% of the original spectral variation. Table 1 presents relative eigenvalues, and estimates and standard error (SE) of relative genetic, permanent environment, and random residual effects. The PCs are sorted according to their contribution to the total spectral variation. The first two PCs explained almost 85% of the total variation (Table 1) which implies that there is strong connection among spectral variables. Variance ratios for the 8 PC traits ranged from 0.011 to 0.285 for genetic effects (heritabilities), from 0.013 to 0.27 for permanent environment effects, and from 0.035 to 0.721 for residual effects (Table 1). Only two of the PCA traits (PC 4 and PC 5) have permanent environment variance ratios higher than that of genetic.

Table 1: Estimates and ±SE of variance ratios for genetic, permanent environment, herd test day and residual random effects.

<table>
<thead>
<tr>
<th>PCs</th>
<th>Relative Genetic (heritability)</th>
<th>Permanent environment</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
</tr>
<tr>
<td>1</td>
<td>58.96</td>
<td>0.154</td>
<td>0.020</td>
</tr>
<tr>
<td>2</td>
<td>25.62</td>
<td>0.264</td>
<td>0.091</td>
</tr>
<tr>
<td>3</td>
<td>7.74</td>
<td>0.146</td>
<td>0.015</td>
</tr>
<tr>
<td>4</td>
<td>3.24</td>
<td>0.057</td>
<td>0.270</td>
</tr>
<tr>
<td>5</td>
<td>1.66</td>
<td>0.011</td>
<td>0.013</td>
</tr>
<tr>
<td>6</td>
<td>1.07</td>
<td>0.229</td>
<td>0.135</td>
</tr>
<tr>
<td>7</td>
<td>0.41</td>
<td>0.211</td>
<td>0.058</td>
</tr>
<tr>
<td>8</td>
<td>0.28</td>
<td>0.285</td>
<td>0.033</td>
</tr>
</tbody>
</table>

Importance of PC traits according to their relative eigenvalues does not agree with their rank according to their heritabilities (Table 1). For instance, PC 8 has the highest heritability but not the largest relative eigenvalue indicating that the largest variation of the spectral data is not of breeding interest. Similar observation could be made for permanent environmental effects.

Estimated variance components of PC traits were back transformed to original spectral variables. Figure 1 presents estimated variance ratios for genetic and permanent environment effects. Variance ratios of the spectral variables ranged from 0.018 to 0.408 for genetic and from 0.002 to 0.184 for permanent environment effects (Figure 1). Implication of the observed genetic variability of individual wave number is difficult to comprehend because milk spectra represent several thousand bio-molecules of the milk (Soyeurt et al., 2010). However, genetic variability of some regions of the FTIR spectra could be interpreted from previous knowledge of the spectra. The fingerprint region of the spectra, approximately between 900 cm\(^{-1}\) and 1500 cm\(^{-1}\), corresponds to C–O and C–C stretching vibration mode (Iñón et al., 2004). In milk, these chemical bonds are...
building blocks of the major milk components such as carbohydrates, lipids and protein. Higher heritabilities observed in this region (Figure 1) could be related with the genetic variability of these milk components. In a similar manner, higher genetic variability observed between 1700 cm\(^{-1}\) – 1800 cm\(^{-1}\) could be related to carbonyl group [C=O] of milk lipids (Iñón et al., 2004), and the heritability peaks between 2800 cm\(^{-1}\) and 3000 cm\(^{-1}\) could be related to C–H group of milk fat.

![Figure 1: Genetic (——) and permanent environment (-----) variance ratios of FTIR spectral traits](image)

Our results confirmed that there is a substantial amount of genetic variation in goat milk FTIR spectra; however, not all FTIR regions are of breeding interest. Further, the method developed here could be useful to monitor goat milk quality due to environment and feeding through prediction of PC traits.

**References**


