Estimating myostatin gene effect on milk performance traits using estimated gene content for a large number of non-genotyped cows

B. Buske¹‡, M. Szydlowski¹,³, C. Verkenne¹ and N. Gengler¹,²

¹University of Liège, Gembloux Agro-Bio Tech, Animal Science Unit, B-5030 Gembloux, Belgium; ²National Fund for Scientific Research, B-1000 Brussels, Belgium; ³Department of Genetics and Animal Breeding, Poznan University of Life Sciences, 60-637 Poznan, Poland

(Received 19 January 2010; Accepted 24 June 2010; First published online 24 August 2010)

The objective of this study was to estimate the myostatin (mh) gene’s effect on milk, protein and fat yield in a large heterogeneous cow population, of which only a small portion was genotyped. For this purpose, a total of 13,992,889 test-day records derived from 799,778 cows were available. The mh gene effect was estimated via BLUP using a multi-lactation, multi-trait random regression test-day model with an additional fixed regression on mh gene content. As only 1,416 animals, (of which 1,183 cows had test-day records) were genotyped, more animals of additional breeds with assumed known genotype were added to estimate the genotype (gene content) of the remaining cows more reliably. This was carried out using the conventional pedigree information between genotyped animals and their non-genotyped relatives. Applying this rule, mean estimated gene content over all cows with test-day records was 0.104, showing that most cows were homozygous 1/1. In contrast, when gene content estimation was only based on genotyped animals, mean estimated gene content over all cows with test-day records was with 1.349 overestimated. Therefore, the applied method for gene content estimation in large populations needs additional genotype assumptions about additional animals representing genetic diversity when the breed composition in the complete population is heterogeneous and only a few animals from predominantly one breed are genotyped. Concerning allele substitution effects for one copy of the ‘mh’ gene variant, significant decreases of $-76.1$ kg milk, $-3.6$ kg fat and $-2.8$ kg protein/lactation were obtained on average when gene content estimation was additionally based on animals with assumed known genotype. Based on this result, knowledge of the mh genotypes and their effects has the potential to improve milk performance traits in cattle.

Keywords: BLUP, estimated gene content, milk performance traits, myostatin gene

Implications
This study describes a method to predict genetic information of single genes for non-genotyped animals via the classical pedigree between genotyped animals and their non-genotyped relatives. The aim is to enlarge the sample of animals for estimating a single gene effect more reliably. The importance lies in the possibility of increasing genetic gain by selecting animals with desirable alleles for single genes that have a major impact on economically important traits.
with genotypic information. Different methods to predict the genotypes or gene content of non-genotyped animals via genotyped animals and the relationships between genotyped and non-genotyped animals were developed (e.g. Van Arendonk et al., 1989; Israel and Weller, 1998). Recently, Gengler et al. (2007) developed a method which has useful advantages in comparison to other methods. This method allows the estimation of gene content of all relatives, including ancestors, for very large pedigrees (even in the presence of pedigree loops and errors), which is not possible when applying some older methods. Another advantage is that this method can include genetic groups. The latter feature allows differentiating allele frequencies in different groups of base animals.

There is an interest in estimating the myostatin (mh) gene effect on milk production traits. The mh gene is responsible for double-muscling in all cattle breeds (e.g. Bellinge et al., 2005), but there are indications that this gene also has an influence on milk performance traits in dairy cattle. However, the effect of the mh gene on milk performance traits is not well known and only a few animals out of a large population are genotyped.

Therefore, the overall objective of this study was to estimate the mh gene effect in a heterogeneous cow population from which only a small part was genotyped. In order to estimate the mh effect more reliably, the sample was extended to include a large number of non-genotyped cows with estimated gene content by using the recently developed method for gene content prediction.

Material and methods

Data structure and genotyping

Data that were used for the routine evaluation of milk production of January 2007 were provided by the Walloon Breeding Association of Belgium. These data included 13 992 889 test-day records (6 138 812 for first, 4 575 525 for second and 3 278 552 for third lactation) for 799 778 dairy and dual purpose (DP) cows in production with 80.2% being purebred, 18.1% crossbreds of two breeds and 1.7% crossbreds of at least three breeds. Breeds for cows in production included mainly Holstein, Belgian-Blanc-Bleu (BBB) and Red-White. The pedigree file contained 1 429 939 animals consisting of 78.1% purebreds with the three main breeds Holstein, BBB and Red-White, 20.5% crossbreds of two breeds and 1.4% crossbreds of at least three breeds. Therefore, breed composition of animals of both the data and pedigree file was fairly heterogeneous but nearly the same. A total of 1416 DP individuals, which included 1183 cows with test-day records, were genotyped for the mh gene according to the method of Fahrenkrug et al. (1999). Genotype frequencies were 0.148 (+/+), 0.334 (mh/mh) and 0.518 (mh/mh) for all animals and 0.153 (+/+), 0.330 (mh/mh) and 0.517 (mh/mh) for cows with test-day records.

Gene content estimation

As the number of genotyped individuals represented only a small proportion of the 1 429 939 animals in the pedigree, animals with ‘assumed’ known genotypes were added to those with known genotypes. For this purpose, the following rules were applied: (i) All BBB animals from the meat type used for artificial insemination and born after 1985 (830 individuals) were assumed to be mh/mh as the mh allele was already considered fixed for this breed type at that time. (ii) All purebred non-BBB animals (659 971 individuals) were assumed to be +/+ as the probability that these animals were carrying at least one mh allele was very small. Therefore, a total of 662 217 individuals were supposed to have a known genotype. For the remaining 767 722 animals, gene content had to be estimated. In order to estimate the gene content of the founders (animals with unknown parents) more precisely, 10 different genetic groups were created according to breed to distinguish BBB from non-BBB animals, herd-book type of the animal (meat or DP) and year of birth. This was performed because of different origins of animals in the pedigree and data file. Gene content for non-genotyped animals was defined as the number of copies of the mh allele and was estimated applying the method described by Gengler et al. (2007). This method predicts gene content for each non-genotyped animal from animals with known genotype using the additive relationship between genotyped and non-genotyped animals. Estimated gene content for non-genotyped animals is therefore an approximation of their true genotype and is a continuous variable varying between 0 for +/+ and 2 for mh/mh animals.

Statistical model to estimate allele-substitution effects

The model for estimating the mh allele effect was based on a multi-lactation, multi-trait (three lactations × three traits) random regression test-day model, which is used for the routine evaluation for milk production in the Walloon Region of Belgium (Auvray and Gengler, 2002). This model was extended for an additional single gene effect and is:

\[ y = Xβ + Z_0 q + \sum_{i=0}^{2} W_i h_i + \sum_{i=0}^{2} Z_i^2 p_i + \sum_{i=0}^{2} Z_i u_i + e \]

where \( y \) is a vector of pre-corrected milk, fat and protein test-day records. Pre-corrections were performed for environmental effects of age considering different rates of maturity for three breed groups (BBB, Holstein, others). The vector \( β \) stands for fixed effects (herd × test-day, fine stage of lactation class, gestation stage and large stage of lactation class × age at calving × season of calving). Thus, overall fixed lactation curves were modelled by fine stage of lactation classes (5 days/class) and larger stage of lactation classes (30 days/class) within age and season of calving. The vector \( h \) stands for herd × period of calving random regression coefficients, \( p \) is a vector of permanent environmental random regression coefficients across lactations, \( u \) is a vector of polygenic random additive regression coefficients and \( e \) is a vector of random residuals. The model used included genetic groups based on breed groups (BBB, Holstein, others), date of birth, sex and region of origin for Holsteins (North America or Europe). The vector \( q \) stands for the known and assumed known gene content for 662 217
production cows, whereas the vector \( \hat{q} \) represents the estimated gene content for the remaining 767 722 production cows. Gene content was not weighted with regard to its derivation (known, assumed known or estimated). The estimated additive fixed regression coefficient on the mh allele effect \( g_1 \) was assumed to be equal across breeds. This simplicity was necessary because a mh gene \( \times \) breed interaction could not be considered due to the assumption that all Holstein cows were +/+ . The matrix \( Z_0 \) is the incidence matrix linking observations to animals, and \( X \), \( W \), \( Z^* \) and \( Z \) are incidence matrices which link observations to further fixed and random effects as described above. The matrix \( Z^* \) stands for the non-zero part of \( Z \). Dominance effects were not considered because pre-investigations using only the 1416 genotyped animals showed that dominance effects were non-significant for all traits. Random regression effects were modelled using modified Legendre polynomials. Subscripts of incident matrices link them to their respective modified Legendre polynomials used, which were:

\[
I_0 = 1 \\
I_1 = \sqrt{3}x \\
I_2 = \sqrt{5/(4(3x^2-1))}
\]

where \( x = -1 + 2 \cdot \text{[days in milk \( -1)/(365 - 1)]} \). (Co)variance components used were those from the routine genetic evaluations obtained previously using EM-REML (Auvray and Gengler, 2002), as described by Gengler et al. (1999). Modified mixed model equation were solved iteratively using a standard preconditioned conjugate gradient solver (e.g. Stranden and Lidauer, 1999).

Standard errors for regression coefficients were estimated using mixed model conjugate gradient normal equations (Harville, 1979). More details about this method are described by Croquet et al. (2006). Significance of allele substitution effects was tested using an approximate \( t \)-test with \( N - \text{rank}(X) \) degrees of freedom, where \( N \) is the number of test-day records and \( X \) is the matrix as described previously.

### Results and discussion

#### Estimated gene content of non-genotyped animals

The evolution of the average gene content for all animals in the pedigree of 1 429 939 animals over the past 40 years depended strongly on its breed and, for BBB animals, on its herd-book type (Table 1). Non-BBB animals were estimated to a near zero value of gene content as expected. BBB animals, which were in their large majority selected since the beginning of the 1970s as a ‘meat type’ showed an expected strong increase of the mh allele. DP-BBB animals starting at a similar level increased slightly and remained then more stable at a gene content of about 1.2. Mean gene content for all cows with test-day records was 0.104, showing that most cows were homozygous ++/+. Such a result is not realistic because it would mean that on average cows with production carried more than one mh allele. There are two reasons for this result. First, the genotypes of the 1416 genotyped animals were only from BBB animals showing a slight tendency in favour of the mh allele. Second, the non-genotyped animals had often weak or non-existing pedigree links to genotyped animals. In this case, the estimated gene content tended towards the genetic group gene content, or, if no information was available to distinguish between groups it tended towards the average gene content. Therefore, the applied assumptions about additional animals in the population led to more realistic results. This result shows that the applied method for gene content estimation is limited, when only a few animals are genotyped as part of a large heterogeneous population.
population which is under selection. The accuracy of gene content estimation for the applied method depends mainly on four factors: First, it increases, when the number of genotyped relatives increases and, second, it increases when the relationship between genotyped and non-genotyped relatives is close. Third, a preferably complete and error-free pedigree is desirable. Errors in the pedigree generally diminish the power of gene content estimation, but the applied method has the property to be fairly robust for hidden errors (Gengler et al., 2007). Fourth, which animals are genotyped is also a relevant concern. In a study by Gengler et al. (2008), gene content was predicted for 1,656,599 non-genotyped cows, from which at least 75% had a genotyped sire or maternal grand-sire or both. A genotyped sire with genotyped cows, from which at least 75% had a genotyped genotype was more informative, that is, carrying the same genotype. Dominance effects, estimated in a pre-investigation of this study with only genotyped cows were found to be non-significant (results not shown). Considering the three lactations separately, a copy of the mh allele led to very highly significant decreases in obtained yields for all traits (Table 3). For all yields, the third lactation was affected more than the first two lactations, contrary to the study by Gengler et al. (2004) where the first lactation was affected the most. These differences can be explained by the low number of known and estimated genotypes in their study in comparison to this study. The power to detect a candidate gene effect depends on the magnitude of the effect that was estimated and its standard error (Israel and Weller, 1998). Although standard errors increased with increasing lactation number due to less records in higher lactation numbers, all estimated mh allele effects were large enough to be significant. However, it should be mentioned that in our study candidate gene effect

**Table 2** Statistics for gene content for cows in production, unweighted and weighted by number of test-day records

<table>
<thead>
<tr>
<th>Number</th>
<th>Gene content (mean)</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cows in production, (unweighted)</td>
<td>799 778</td>
<td>0.104</td>
</tr>
<tr>
<td>Weighted by number of test-day records</td>
<td>13,992,889</td>
<td>0.088</td>
</tr>
</tbody>
</table>

1Gene content varied from 0 (+/+ to 2 mh/mh) and are continuous values for non-genotyped cows.

**Table 3** Additive effects\(^1\) and s.e. in kg for each lactation (305 days) for milk, fat and protein yield for all 799 778 cows in production with known and estimated genotype

<table>
<thead>
<tr>
<th>Lactation</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait</td>
<td>Effect</td>
<td>s.e.</td>
<td>t-value</td>
</tr>
<tr>
<td>Milk</td>
<td>−70.80</td>
<td>8.30</td>
<td>8.53***</td>
</tr>
<tr>
<td>Fat</td>
<td>−3.02</td>
<td>0.35</td>
<td>8.62***</td>
</tr>
<tr>
<td>Protein</td>
<td>−2.57</td>
<td>0.25</td>
<td>10.28***</td>
</tr>
</tbody>
</table>

\(^1\)Allele substitution effect for one copy of the ‘mh’ allele. 
***P < 0.0001.

**Additive effects of the myostatin gene**

Cows carrying the mh allele produced less milk and had lower fat and protein yields. Allele substitution effects for one copy of the mh allele through all lactations were −76.06, −3.62 and −2.84 kg/lactation period (305 days) for milk, fat and protein yield, respectively. These values were lower than those found by Buske et al. (2010) in a previous study using a Bayesian approach for only 1455 genotyped BBB cows. This difference might be explained by assumptions of this study, which included that allele substitution effects were the same across several breeds, which may not be accurate. Evidence of this includes the relatively high standard errors for the mh gene effect estimation found in this study. Therefore, the necessity to include a large number of animals from another breed to estimate gene content more reliably has the disadvantage that only an average allele substitution effect across breeds can be estimated, when those animals are assumed being un-informative, that is, carrying the same genotype. Dominance effects, estimated in a pre-investigation of this study with only genotyped cows were found to be non-significant (results not shown). Considering the three lactations separately, a copy of the mh allele led to very highly significant decreases in obtained yields for all traits (Table 3). For all yields, the third lactation was affected more than the first two lactations, contrary to the study by Gengler et al. (2004) where the first lactation was affected the most. These differences can be explained by the low number of known and estimated genotypes in their study in comparison to this study. The power to detect a candidate gene effect depends on the magnitude of the effect that was estimated and its standard error (Israel and Weller, 1998). Although standard errors increased with increasing lactation number due to less records in higher lactation numbers, all estimated mh allele effects were large enough to be significant. However, it should be mentioned that in our study candidate gene effect
estimation could be biased due to ignorance as to whether or not animals were selected randomly for genotyping and because of the inclusion of a large number of animals with assumed genotypes.

Conclusions

Results showed that in order to estimate the gene content of a large population, the inclusion of additional (e.g. genotypic) assumptions such as knowledge about phenotypic selection or genetic groups for non-genotyped animals is necessary when genotyped animals make up only a small part of a large and heterogeneous population that is under selection. Otherwise, a reliable genotype estimation remains difficult. This is more pronounced when only one gene is considered and no information about linkage to neighbouring genotyped loci can be exploited. Hence, further genotyping of important animals (e.g. sires with many daughters in production) will be necessary to estimate the gene content for single genes more reliable. Concerning mh allele substitution effects, one copy of the mh allele led to significant decreases in milk, fat and protein yield. Therefore, knowledge of mh genotypes has the potential to improve milk performance traits in cattle. This is also interesting for countries, in which the mh status for their local breeds is unknown and improvements due to feeding and management systems are limited.

Acknowledgements

Nicolas Gengler, who is Research Associate of the National Fund for Scientific Research (Brussels, Belgium), acknowledges his support. The authors gratefully acknowledge the financial support of the Ministry of Agriculture of the Walloon Region of Belgium (MRW-DGA; Namur, Belgium) projects D31-1112 and D31-1168. Additional support was provided through grants 2.4507.02F (2) and F.4552.05 of the National Fund for Scientific Research. The authors wish to thank S. Loker, University of Guelph (Canada) for reviewing the manuscript.

Estimating myostatin gene effect on milk performance traits

References