ABSTRACT

High-throughput cow genotyping has opened new perspectives for genome-wide association studies (GWAS). Directly recorded phenotypes and several records per animal could be used. In this study, a GWAS on lactation curve traits of 337 Italian Simmental cows genotyped with the Illumina (San Diego, CA) low-density BeadChip (7K) was carried out. Scores of the first 2 principal components extracted from test-day records (7 for each lactation) for milk yield, fat and protein percentages, and somatic cell score were used as phenotypes. The first component described the average level of the lactation curve, whereas the second summarized its shape. Data were analyzed with a mixed linear model that included fixed effects of herd, calving month, calving year, parity, SNP genotype, and random effects of animal and permanent environment. All statistically significant markers (Bonferroni corrected) were associated with the average level component (2 for milk yield, 9 for fat percentage, 6 for protein percentages, and 1 for somatic cell score). No markers were found to be associated with the lactation curve shape. Gene discovery was performed using windows of variable size, according to the linkage disequilibrium level of the specific genomic region. Several suggestive candidate genes were identified, some of which already reported to be associated with dairy traits, such as DGAT1. Others were involved in lipid metabolism, in protein synthesis, in the immune response, in cellular processes, and in early development. The large number of genes flagged in the present study suggests interesting perspectives for the use of low-density genotyped females for GWAS, also for novel phenotypes that are not currently considered as breeding goals.

Key words: lactation curve, principal component analysis, genome-wide association study, low-density panel

INTRODUCTION

High-throughput SNP platforms have been used in several genome-wide association studies (GWAS) in dairy cattle. These have been often carried out within genomic selection (GS) programs, where bulls are preferentially genotyped. However, the need to enlarge the size of the reference populations has led to genotyping cows. In most cases, females are genotyped with low-density SNP panels that yield less accurate genomic evaluations but at approximately half the price (Wiggans et al., 2013).

The use of cow data in GS programs has raised some issues. For example, phenotypes need to be adjusted to be comparable with those of bulls (Wiggans et al., 2011). However, advantages in terms of direct genomic value accuracy and in the economic returns of GS schemes that include females have been underlined in studies carried out both on simulated and real data (Lourenco et al., 2014; Thomasen et al., 2014).

So far, polygenic EBV, daughter yield deviations, or deregressed proofs for traits of interest have been the most frequently used dependent variables in GWAS for dairy traits. The use of genotyped females allows the direct modeling of recorded phenotypes. Moreover, multiple records are often available for each cow (e.g., multiple measures per lactation, successive lactations). An appealing, straightforward consequence of the use of repeated records is that SNP effects can be fitted along the whole lactation curve using longitudinal models (Strucken et al., 2015).

Candidate gene effects on lactation curves for milk production traits have been investigated using mathematical functions (Strucken et al., 2011; Szyda et al., 2014). In particular, values of function parameters or coefficients of orthogonal polynomials estimated for individual lactation curves were used as dependent variables. A GWAS for different measures of lactation persistency from random regression models was performed on primiparous Holstein and Jersey cows (Pryce et al., 2010). Genomic regions associated with lactation persistency in both breeds were detected on BTA6 and
BTA26. A study carried out on German Holstein cows divergently selected for milk yield used Wilmink function parameters as phenotypes (Strucken et al., 2012). Single nucleotide polymorphisms associated with lactation curve traits were detected, especially for lactation persistency. Moreover, a dependence of results on parity order was highlighted.

Actually, modeling individual lactation curves is hampered by the large variability between cows (Olori et al., 1999). Individual differences are further enhanced by mathematical artifacts (e.g., negative or very high predicted values of test-day yields at the beginning and at the end of the lactation trajectory). These problems often occur when models are fitted to patterns that markedly differ from the shape of the standard lactation curve (Macciotta et al., 2011). The resulting huge variability of parameter estimates, both in magnitude and sign, suggests a great care when these are used in further analyses (Macciotta et al., 2005). As an alternative, model-free algorithms able to derive measures of lactation curve traits without specific assumptions on data structure could be used. An example is principal component analysis (PCA). A PCA carried out on test-day records for milk yield treated as different traits yielded 2 transformed variables related to (1) the average production over the entire lactation (LVL) and (2) the shape of the lactation curve (SHP; Macciotta et al., 2006); both traits had moderate values of heritability.

The aim of the present work was to perform a GWAS on lactation curve traits using model-free PCA; dependent variables were obtained from the PCA of test-day records for milk production traits. Cows were genotyped with a low-density (7K) SNP panel, as is common in GS programs. All animals belonged to the Italian Simmental breed, the third-ranked breed in Italy for milk production (42,133 recorded lactations in 2013) after Italian Holstein and Italian Brown (ICAR 2014).

### MATERIALS AND METHODS

Data consisted of 10,605 test-day records for milk production and milk composition traits (milk yield, fat and protein percentages, and SCS). Records were from 1,515 lactations of 337 Italian Simmental cows. Animals were farmed in 120 herds. Seven test-day (TD) records were considered for each lactation. Extra TD for lactations with more than 7 records were deleted. All cows were genotyped with the 7K Illumina BeadChip (Illumina, San Diego, CA). Marker edits were on call rate (>0.99) and minor allele frequency (>0.01). After edits, 6,891 markers were retained for the analysis.

Principal component analysis was performed on the 7 TD records, separately for each lactation. Scores of the first 2 principal components (PC1 and PC2) for each lactation were then calculated by multiplying the row vector of the standardized original variables by the column vector of the corresponding eigenvector coefficients. Scores were used as dependent variables in the association study carried out with the following mixed linear model:

\[
y_{ijklmnop} = \mu + H_i + M_j + Y_k + Par_l + SNP_m + a_n + p_o + e_{ijklmnop},
\]

where \( y \) = principal component score; \( \mu \) = overall mean; \( H \) = fixed effect of the ith herd; \( M \) = fixed effect of the jth calving month (12 mo); \( Y \) = fixed effect of the kth calving year (from 2002 to 2013); \( Par \) = fixed effect of lth parity (from 1 to 6, > 6); \( SNP \) = fixed covariable of the mth SNP marker genotype (coded as 0, 1, 2 according to the copies of the second allele); \( a \) = random additive effect of the nth animal; \( p \) = random permanent environmental effect of the oth lactation; and \( e \) = random residual. The animal and permanent environment random effects were assumed to be normally distributed as \( N(0, \sigma^2_a) \) and \( N(0, \sigma^2_p) \), respectively, where \( A \) is the pedigree relationship matrix, \( I \) is an identity matrix, and \( \sigma^2_a \) and \( \sigma^2_p \) are additive genetic and permanent environmental variances, respectively. The random residual was assumed to be normally distributed as \( N(0, \sigma^2_e) \), where \( \sigma^2_e \) is the residual variance.

The mixed model was solved using a REML algorithm implemented in the ASREML software (Gilmour et al., 2000). Values published by Macciotta et al. (2006) were used as variance priors (PC1 = 0.4427, 0.4654, and 1.4005; PC2 = 0.05264, 0.02954, and 0.66050, for \( \sigma^2_a, \sigma^2_p, \) and \( \sigma^2_e \), respectively).

Bonferroni-corrected significance levels for the SNP effects were calculated to account for multiple testing: uncorrected \( P \)-values were multiplied by the number of tests performed (i.e., 6,891). The SNP were considered significantly associated with the considered trait when the corrected \( P \)-value was lower than 0.05.

Gene discovery was performed in the genomic regions located around the significant SNP. The width of these intervals was based on the linkage disequilibrium (LD) of the region. The LD was calculated on a sample of 479 Italian Simmental bulls genotyped with the 54K Illumina BeadChip in a previous study (Pintus et al., 2012). For each significant SNP detected in the present study, the value of the squared coefficient (\( r^2 \)) statistic with all other SNP located in the same chromosome
was calculated using the Haploview software (Barrett et al., 2005). Then the distance between the significant SNP and the farthest SNP having an $r^2 > 0.15$ was calculated. A window was then defined by considering such distance both upstream and downstream the position of the significant SNP. Annotated genes located in the windows were derived from UCSC Genome Browser Gateway (http://genome.ucsc.edu/). The SNP and gene positions were obtained from the UMD3.1 Bovine genome assembly (Zimin et al., 2009).

### RESULTS

**PCA**

The first 2 principal components extracted from the correlation matrix of the TD records accounted for most of the original variance. The sum of the first 2 eigenvalues was 83% for milk yield, 50% for fat percentage, 57% for protein percentage, and 42% for SCS (Table 1). The structure of eigenvectors showed an association between PC1 and all the TD records. Principal component 2 was positively associated with the first and negatively with the second part of lactation, respectively. Such a structure could be observed for all the 4 considered traits. The first eigenvalue was always markedly larger than the second, in agreement with previous studies (Macciotta et al., 2006). Differences in magnitude can be observed between traits (Table 1).

Figures 1 to 4 report average lactation patterns for animals grouped according to PC1 and PC2 scores for the 4 considered traits. In particular, Figures 1a and b highlight the role of the 2 principal components as phenotypic indices of level of production (LVL) and lactation curve shape (SHP) for milk yield, respectively. The first 2 principal components have a similar meaning also for fat percentage (Figures 2a and 2b), protein percentage (Figures 3a and 3b), and SCS (Figures 4a and 4b).

### Table 1. Eigenvectors and eigenvalues of the first 2 principal components (PC1 and PC2) extracted for the 7 test-day records for milk yield, fat and protein percentage, and SCS

<table>
<thead>
<tr>
<th>Test day</th>
<th>Milk yield</th>
<th>Fat</th>
<th>Protein</th>
<th>SCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PC1 PC2</td>
<td>PC1 PC2</td>
<td>PC1 PC2</td>
<td>PC1 PC2</td>
</tr>
<tr>
<td>Test 1</td>
<td>0.32 0.60</td>
<td>0.19 0.74</td>
<td>0.14 0.75</td>
<td>0.26 0.78</td>
</tr>
<tr>
<td>Test 2</td>
<td>0.38 0.38</td>
<td>0.32 0.49</td>
<td>0.34 0.42</td>
<td>0.32 0.48</td>
</tr>
<tr>
<td>Test 3</td>
<td>0.40 0.16</td>
<td>0.38 0.11</td>
<td>0.35 0.19</td>
<td>0.38 0.17</td>
</tr>
<tr>
<td>Test 4</td>
<td>0.40 0.02</td>
<td>0.40 −0.07</td>
<td>0.47 −0.01</td>
<td>0.43 −0.22</td>
</tr>
<tr>
<td>Test 5</td>
<td>0.40 −0.20</td>
<td>0.44 −0.19</td>
<td>0.43 −0.14</td>
<td>0.44 −0.21</td>
</tr>
<tr>
<td>Test 6</td>
<td>0.38 −0.38</td>
<td>0.44 −0.30</td>
<td>0.44 −0.28</td>
<td>0.39 −0.11</td>
</tr>
<tr>
<td>Test 7</td>
<td>0.34 −0.54</td>
<td>0.41 0.24</td>
<td>0.35 −0.37</td>
<td>0.38 −0.17</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>0.71 0.12</td>
<td>0.35 0.15</td>
<td>0.42 0.15</td>
<td>0.39 0.13</td>
</tr>
</tbody>
</table>

**Figure 1.** Average lactation curves for milk yield of groups of cows of different principal component (PC) 1 (a) and PC2 (b) score classes ($\bigstar = -2; \blacksquare = -2$ to $-1; \blacktriangle = -1$ to $0; \blacklozenge = 0$ to $1$; continuous line $= 1$ to $2$; dotted line $> 2$). Points are plotted for the average day in milk on each test day.
Association Study

Eighteen SNP significantly associated (Bonferroni-corrected \( P < 0.05 \)) to the principal component scores for the considered traits (Table 2) were detected. Most of them were for fat and protein percentage (9 and 6, respectively). All significant SNP were associated with PC1 (i.e., the variable that expressed the level at which the lactation curve is located). No SNP were found to be significantly associated with lactation curve shape (i.e., PC2).

**Milk Yield.** The top significant SNP for milk yield LVL (corrected \( P = 0.003 \)) was located on BTA6 at approximately 89 Mb. A strong association between this region and clinical mastitis, milk yield, and protein percentage was reported in dairy cattle (Sahana et al., 2014). Some interesting genes are located in the interval of approximately 0.48 Mb calculated around the significant SNP (Table 2). One is the \( GC \) (groups-specific component), a gene that encodes for a vitamin D-binding protein. It is involved in several physiological functions, such as the modulation of inflammatory and immune response, binding of FA, and bone development (Speeckaert et al., 2014), and has been suggested as a putative candidate for milk yield (Raven et al., 2014) and clinical mastitis (Sahana et al., 2014). The
The CN cluster is also located in this genomic region, but it did not fall in the considered interval. The second SNP in the rank mapped on BTA21, close to the \textit{NRTK3} (neurotrophic tyrosine kinase, receptor, type 3) locus (Table 2). This gene encodes for a membrane protein receptor and it is involved in the determinism of a type of breast carcinoma in humans (Tognon et al., 2002). Other interesting genes located in this region are 2 mitochondrial ribosomal proteins (\textit{MRLP1} and \textit{MRPS11}) that have been reported as candidates for mitochondrial disorders in humans (Vasta et al., 2009).

**Fat Percentage.** The most significant SNP ($P < 0.001$) associated with the fat content LVL was located on BTA23 (Figure 5). A suggestive candidate gene that maps in the window around this marker is the desmosplakin locus. This gene encodes for a protein involved in the structure of desmosomes, intercellular junctions that provide tissue integrity at epithelial level (Gar-

![Figure 4. Average lactation curves for SCS of groups of cows of different principal component (PC) 1 (a) and PC2 (b) score classes (• = −2; ■ = −2 to −1; ▲ = −1 to 0; ○ = 0 to 1; continuous line = 1 to 2; dotted line = >2). Points are plotted for the average day in milk on each test day.](image)

### Table 2. Markers significantly associated (Bonferroni-adjusted level of significance <0.05) with scores of principal components representing lactation curve traits

<table>
<thead>
<tr>
<th>Marker</th>
<th>Bos taurus autosome</th>
<th>Position (bp)</th>
<th>Trait$^1$</th>
<th>$P$-value</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTB-01654826</td>
<td>6</td>
<td>88,891,318</td>
<td>PC1 MY</td>
<td>0.003</td>
<td>±242,276</td>
</tr>
<tr>
<td>Hapmap38505-BTA-51760</td>
<td>21</td>
<td>19,193,100</td>
<td>PC1 MY</td>
<td>0.015</td>
<td>±690,747</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-11659</td>
<td>23</td>
<td>47,677,534</td>
<td>PC1 FP</td>
<td>0.0003</td>
<td>±5,663</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-75852</td>
<td>7</td>
<td>71,545,383</td>
<td>PC1 FP</td>
<td>0.001</td>
<td>±214,702</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-34135</td>
<td>14</td>
<td>1,675,278</td>
<td>PC1 FP</td>
<td>0.003</td>
<td>±783,698</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-18926</td>
<td>3</td>
<td>99,592,696</td>
<td>PC1 FP</td>
<td>0.009</td>
<td>±2,100,054</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-24012</td>
<td>17</td>
<td>74,948,921</td>
<td>PC1 FP</td>
<td>0.016</td>
<td>±267,855</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-4939</td>
<td>14</td>
<td>1,801,116</td>
<td>PC1 FP</td>
<td>0.029</td>
<td>±2,574,789</td>
</tr>
<tr>
<td>Hapmap34329-BES11_Contig247_1378</td>
<td>2</td>
<td>98,446,391</td>
<td>PC1 FP</td>
<td>0.031</td>
<td>±132,872</td>
</tr>
<tr>
<td>Hapmap41022-BTA-55560</td>
<td>23</td>
<td>15,147,471</td>
<td>PC1 FP</td>
<td>0.044</td>
<td>±1,729,322</td>
</tr>
<tr>
<td>Hapmap55987-ss462997</td>
<td>19</td>
<td>24,972,085</td>
<td>PC1 FP</td>
<td>0.05</td>
<td>±256,339</td>
</tr>
<tr>
<td>BTB-01223907</td>
<td>16</td>
<td>32,653,232</td>
<td>PC1 PP</td>
<td>0.0205</td>
<td>±10,322,590</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-55674</td>
<td>12</td>
<td>16,069,827</td>
<td>PC1 PP</td>
<td>0.0215</td>
<td>±373,760</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-97136</td>
<td>6</td>
<td>45,909,053</td>
<td>PC1 PP</td>
<td>0.0240</td>
<td>±868,964</td>
</tr>
<tr>
<td>UA-IFASA-8256</td>
<td>7</td>
<td>21,704,630</td>
<td>PC1 PP</td>
<td>0.0341</td>
<td>±1,619,254</td>
</tr>
<tr>
<td>BTA-102818-no-rs</td>
<td>13</td>
<td>83,632,355</td>
<td>PC1 PP</td>
<td>0.0423</td>
<td>±244,086</td>
</tr>
<tr>
<td>Hapmap50096-BTA-75536</td>
<td>5</td>
<td>11,284,482</td>
<td>PC1 PP</td>
<td>0.0492</td>
<td>±28,293</td>
</tr>
<tr>
<td>ARS-BFGL-NGS-84222</td>
<td>22</td>
<td>531,301</td>
<td>PC1 SCS</td>
<td>0.0360</td>
<td>±43,004</td>
</tr>
</tbody>
</table>

$^1$Trait = PC1MY (first principal component extracted from milk yield test data); PC1FP (first principal component extracted from fat percentage test data); PC1PP (first principal component extracted from protein percentage test data); PC1SCS (first principal component extracted from SCS test data).
The second significant marker was located on BTA7 at approximately 71.5 Mb. The clathrin interactor 1 (\textit{CLINT1}) locus that encodes for a protein which is involved in the vesicle trafficking (Dodd et al., 2009) maps in this region. This gene has been found to be associated with skin color in chicken (Sun et al., 2013).

Two significant SNP for LVL of fat content were highlighted on BTA14. They both map in a very dense region where the \textit{DGAT1} is located. This gene has a well-known major effect on milk fat in cattle (Grisart et al., 2002). The first marker (ARS-BFG-NGS-34135) was reported to be statistically associated with milk yield and fat and protein percentages in a large multibreed study (Raven et al., 2014). The zinc finger protein 34 (\textit{ZNF 34}) and the glutamic pyruvate transaminase (\textit{GPT}) are also located in the interval around this SNP. Associations between these genes and fat yield and percentage have been detected in Chinese Holstein (Jiang et al., 2014). Moreover, the reductase gene family has been reported under balancing selection in a recent comparison between \textit{Bos taurus} and \textit{Bos indicus} genomes (Porto-Neto et al., 2013). The marker found on BTA17 (Table 2) defined a 0.53-Mb window where claudin 5 (\textit{CLDN5}) is mapped. This gene encodes a membrane protein that is a component of the tight junctions. The significant SNP on BTA2 was found in a region that harbors several interesting genes. One is the long-chain acyl CoA dehydrogenase (\textit{ACAD}) gene, which encodes for a key enzyme of the FA metabolism in the liver (Li et al., 2013). Another gene located in this genomic region is the myosin light chain 1 (\textit{MYL1}), whose expression has been related to the physiological status (lactation vs. puberty) in cattle (Ron et al., 2007).

The second significant SNP found on BTA23 was located in a region that harbors some genes of potential interest. One is apolipoprotein B mRNA-editing enzyme...
catalytic subunit 2 (APOBEC2), which is expressed in the muscle and affects muscle fiber ratio and body mass in mice (Sato et al., 2010). Another is the MyoD family inhibitor (MDFI) that was proposed as a suggestive candidate gene for a QTL that affects fatness traits in pigs (Huang et al., 2011). Finally, also the interval around the significant marker found on BTA19 included different genes of potential interest. The ATP2A3 that is involved in calcium mobilization in the cell, could be mentioned. This gene has been recently reported as a selection signature in Ethiopian cattle populations (Edea et al., 2014). Another interesting gene is the olfactory receptor, family 1, subfamily E, member 2-like (LOC618124). Genes of the olfactory receptor family have been found in selection signatures in cattle (Qanbari et al., 2010).

**Protein Percentage.** The most significant SNP for protein percentage was located on BTA16 (Figure 6). This marker showed the largest LD (Table 2). Among the genes that map in the surrounding interval, the EFCAB2 is of potential interest. It is involved in the microarchitecture of the bone in humans (Mohan et al., 2013). Another interesting gene is the SYMD3, which contributes to neuromuscular processes and has been found in a region of deleted copy number variation (CNV) in Korean cattle (Shin et al., 2014). The second marker in order of importance for protein percentage was located on BTA12 at approximately 16 Mb. In this region is the spermatid associated (SPERT), which has been reported in a region of selection signature in German Holsteins (Qanbari et al., 2010). Another interesting gene located close to the significant marker is the lymphocyte cytosolic protein 1 (LCP1), a highly conserved protein of the cattle genome (Lemay et al., 2009). The expression of this gene in the mammary tissue has been found to be associated with infections by *Staphylococcus aureus* (Huang et al., 2014).

The third marker was detected on BTA6. It is placed close to a gene cluster that included the leucine-rich repeat LGI family, member 2 (LGI2) and the DEAH (Asp-Glu-Ala-His) box polypeptide 15 (DHX15), involved in the immune response mechanism. Associations between these 2 genes and fat and protein percentages in German Holstein and Fleckvieh have been reported, respectively (Weikard et al., 2012). Another interesting gene located in this region is the superoxide dismutase 3 (SOD3). It has been found to be associated with intake in beef cattle (Al-Husseini et al., 2014). The fourth marker affecting the PCI for protein percentage was located on BTA7, in a region characterized by a relevant LD (Table 2). Several interesting genes could be found in the calculated window. Of particular interest is the eukaryotic translation elongation factor 2 (EEF2), involved in milk protein synthesis in the mammary gland as a mediator of the effect of growth hormone (Hayashi...
et al., 2009). Moreover, it has been found to be downregulated in cattle experimentally infected by bovine tuberculosis (Meade et al., 2007). Another gene of interest in this region is the integrin β1-binding protein 3 (ITGB1BP3), which was found under positive selection in a comparison between Jersey, Guernsey, and Zebuine breeds (Porto-Neto et al., 2013). Intervals surrounding the last 2 markers for protein percentage, located on BTA13 and BTA5 respectively, did not include any annotated gene for the bovine genome.

**SCS.** The only significant marker for SCS was located at the beginning of BTA22. In this region, the vesicular, overexpressed in cancer, prosurvival protein 1 (VOPP1) maps. This gene encodes for a protein that is involved in cellular apoptosis in vertebrates (Pei and Grishin, 2012).

**DISCUSSION**

**PCA**

The PCA performed on test-day records for milk production traits was able to synthesize the main aspects of the lactation pattern (i.e., the general level of production and its shape). The use of PCA sometimes results in a first component correlated with almost all original variables that absorbs most of the total variance. In these cases, the eigenvectors of the other extracted PC often show some variations (values that increase or decrease, or that change sign) in comparison with PC1 (Stearns et al., 2005). These eigenvector structures allow inference of specific aspects of relationships between groups of variables that cannot be seen in the first PC (Jombart et al., 2009). In the case of the lactation curve, the predominance of the first eigenvalue over the second may be also an expression of the larger heritability of milk yield compared with lactation persistency (Cole and VanRaden, 2006; Cole and Null, 2009). The interpretation of eigenvalues as expression of the genetic contribution of principal components to phenotypes was also proposed by Kirkpatrick and Meyer (2004).

Also of interest is the difference that can be observed between the eigenvalues of the first principal component in the 4 traits (Table 1). The largest value was for milk yield, a trait characterized by a polygenic background. The smallest eigenvalue was for fat percentage, a trait that is genetically determined by few genes with large effects (Hayes et al., 2010). From these figures, a relationship between magnitude of the eigenvalue of PC1 and the genetic determinism of the trait could be inferred.

Principal component analysis could, therefore, be seen not only as a dimension-reduction technique, but also as an approach for investigating the genetic determinism of traits. Studies carried out both on dairy and meat traits underlined a higher efficiency of PCA in comparison with univariate analyses for the detection of SNP associations and of QTL with pleiotropic effects (Stearns et al. 2005; Bolormaa et al., 2010).

**Association Analysis**

A relevant number of significant associations were detected in spite of the low-density marker map used. Such a result may be somewhat unexpected because the extent of LD between marker and QTL is one of the main factors affecting the power of GWAS (Powell et al., 2012). Conversely, a large number of significant markers were reported in a previous research carried out on females (Cole et al., 2011). Furthermore, some of the most significant SNP markers found in our study confirmed previous reports on dairy cattle (Cole et al., 2011; Raven et al., 2014; Wang et al., 2014).

It is worth remembering that GWAS on males are usually carried out on progeny tested or on AI bulls (i.e., on the top animals of the breed). A reduction of the genetic variability in comparison with females could be therefore expected. Conversely, it has to be remembered that the Italian Simmental is a dual-purpose breed. Bulls are selected by combining performance test for beef traits and progeny test for dairy traits respectively. Moreover, limitation exists in the number of semen doses that can be used per bull (7,000). These 2 aspects may have contributed to reduce the selection pressure and to maintain genetic variation.

The kind of analyzed traits could be a further cause of results here obtained. Strucken et al. (2011) found that the use of lactation curve parameters instead of yield data as dependent variables provided greater power in detecting associations. Authors explained their results with an increase of the genetic variance of the considered trait. In the present study, the Bonferroni correction of the SNP significance level was implemented to prevent the occurrence of false positives. Moreover, population stratification was accounted for by including the polygenic effect in the model.

All significant associations were found only for the principal component that described the level at which the lactation curve is located. No markers were detected for the shape of the lactation curve. This result is in contrast with previous studies that detected SNP associated with lactation persistency. Some points could be discussed to find possible reasons for these discrepancies, first of all the kind of dependent variable used. There is a lack of consensus on a suitable measure of lactation persistency and this trait has been defined in many ways. The measure of lactation curve shape used...
in the present paper is completely uncorrelated from milk yield (i.e., \( r = 0 \) for PC1 and PC2), and it explains a reduced quota of variation compared with the LVL component. This may be also an explanation of the absence of common significant markers across traits. Conversely, in the paper by Strucken et al. (2011), persistency is defined by using the parameter \( c \) of the Wilmink function, which is usually highly correlated with the parameter \( a \) (i.e., the one related to yield; Macciotta et al., 2005). A second point is represented by the effect of parity. It is widely acknowledged that first-calving cows have a markedly flatter curve in comparison with older parities. In the paper of Pryce et al. (2010), only primiparous cows were considered; moreover, Strucken et al. (2011) found significant associations only when first lactations were analyzed separately from those of older animals. In the present paper, cows of different parity were included in the data set and primiparous represented about 25% of the records. Finally, it should be remembered that variation of persistency explained by significant SNP in previous researches was rather low, 1 to 2% in the study of Pryce et al. (2010).

**Gene Discovery**

The intervals surrounding the significant SNP allowed for the detection of several suggestive candidate genes. Most of them were found for milk fat percentage. Some of these genes were already known to affect dairy traits, such as *DGAT1*. This is a rather common result for GWAS, but it was somewhat unexpected for Italian Simmental. Pintus et al. (2012) hypothesized that the low direct genomic value accuracy for fat percentage obtained in Italian Simmental bulls was due to the absence of allele segregation at the *DGAT1* locus in this breed. The frequency of the favorable *DGAT1* allele was >0.99 in a previous work carried out on 95 cows (Scotti et al., 2010). Results of the present study have been obtained on larger sample. Moreover, the 2 significant markers found in the region of BTA14 that harbors the *DGAT1* locus have been reported also for other breeds (Cole et al., 2011; Raven et al., 2014; Wang et al., 2014). A recent study carried out on Italian Simmental bulls reported a SNP in the promoter region of the *DGAT1*, but no associations with dairy traits were found (Chessa et al., 2015). In any case, it is worth remembering that several genes map in this genomic region; thus, SNP significance could be also due to the effect of other genes.

Most of identified putative candidate genes have a biological connection with the lactation. Examples are those involved in the lipid metabolism (*ACAD*), in protein synthesis (*EEF2*), in the integrity of the epithelium (*CLDN5* and the desmoplakin), and in the immune response (*DHX15, GC*, and *LCP1*). Of interest is also the gene found on BTA6 that encodes antioxidant enzyme *SOD3*, whose expression has been found to be down-regulated in the mammary tissue of cows fed a diet rich in polyunsaturated acid (Côrtes et al., 2012). This gene is involved in the determinism of feed intake in cattle (Al-Husseini et al., 2014). Other genes, which map in genomic regions where selection signatures have been detected, are of more general functions. Examples are olfactory receptors, transmembrane proteins involved cellular processes, and mediators of early development as *LOC618124, VOPP1*, and *ITGB1BP3*, respectively.

**CONCLUSIONS**

The GWAS carried out in Italian Simmental cows highlighted some markers statistically associated with lactation curve traits. In particular, significant associations were found for the principal component that describes the production level at which the lactation curve for different dairy traits is located. No significant SNP were found for lactation curve shape. This result disagrees with previous studies on dairy cattle that report significant associations between genomic regions and lactation persistency. Reasons can be found in the kind of measure used, in the dual-purpose aptitude of the Italian Simmental breed, and in the composition of the sample of animals considered (especially concerning parity). Conversely, some putative candidate genes detected in the present study were found to be associated with production traits in previous research.

In spite of the low-density marker map used, a relatively large number of significant markers was detected. These results suggest the use of low-density genotyped females for GWAS and for novel phenotypes (e.g., milk FA spectrum, milk coagulation properties) that are not currently measured in breeding programs.

**ACKNOWLEDGMENTS**

This research has been funded by the Italian Ministry of Agriculture (Rome, Italy; Grant INNOVAGEN, Project Leader Alessio Valentini).

**REFERENCES**


