

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1687015> since 2019-02-05T15:58:44Z

*Published version:*

DOI:10.3168/jds.2015-9500

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

This is the author's final version of the contribution published as:

Macciotta NPP; Gaspa G; Bomba L; Vicario D; Dimauro C; Cellesi M; Ajmone-Marsan P.

Genome-wide association analysis in Italian Simmental cows for lactation curve traits  
using a low-density (7K) SNP panel

JOURNAL OF DAIRY SCIENCE, 98:8175:8185, 2015,  
doi: 10.3168/jds.2015-9500

The publisher's version is available at:

<https://www.sciencedirect.com/science/article/pii/S0022030215006669>

When citing, please refer to the published version.

Link to this full text:

<http://hdl.handle.net/2318/1687015>

This full text was downloaded from iris-AperTO: <https://iris.unito.it/>

1  
2  
3 **GWAS FOR LACTATION CURVE TRAITS**  
4

5 **Interpretive summary**  
6

7 **Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a**  
8 **low density (7K) SNP panel.**  
9

10 by Macciotta et al.

11 A genome wide association study on PCA-derived lactation curve traits was performed on a sample  
12 of Italian Simmental cows genotyped with a low density SNP panel. Eighteen significant SNP were  
13 detected. Gene discovery highlighted some interesting candidate genes Results suggest interesting  
14 perspectives for the use of low density genotyped females for GWAS,.  
15

## GWAS FOR LACTATION CURVE TRAITS

### Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low density (7K) SNP panel

N. P. P. Macciotta<sup>\*1</sup>, G. Gaspa<sup>\*</sup>, L. Bomba<sup>‡</sup>, D. Vicario<sup>†</sup>, C. Dimauro<sup>\*</sup>, M. Cellesi<sup>\*</sup>, and P. Ajmone-Marsan<sup>‡</sup>

<sup>\*</sup> Dipartimento di Agraria, Università di Sassari, Sassari, Italy,

<sup>†</sup> ANAPRI, Udine, Italy,, Università di Sassari,

<sup>‡</sup> Università Cattolica del Sacro Cuore, Piacenza, Italy

<sup>1</sup>Corresponding author: Nicolò P.P. Macciotta, Dipartimento di Agraria, Sezione Scienze

Zootecniche, Università di Sassari, viale Italia 39, 07100 Sassari, Italy. Phone number: 0039

079229299. Fax number: 0039 079229302. e-mail: macciott@uniss.it

#### 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 low density beadchip (7K) was carried out. Scores of the first two principal components extracted from test day records (seven for each lactation) for milk yield, fat and protein percentages, and somatic cell score (SCS) were used as phenotypes. The first component described the average level (LEVEL) of the lactation curve, whereas the second summarized its shape (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  $P < 0.05$ ) were associated to the LEVEL component (two

44 for milk yield, nine for fat percentage, six for protein percentages and one for SCS, respectively).  
45 No markers were found to be associated to the lactation curve shape. Gene discovery was  
46 performed using windows of variable size, according to the linkage disequilibrium level of the  
47 specific genomic region. Several suggestive candidate genes were indentified, some of which  
48 already reported to be associated with dairy traits as *DGATI*. Others were involved in lipid  
49 metabolism, in protein synthesis, in the immune response, in cellular processes, and in early  
50 development. The large number of genes flagged in the present study suggests interesting  
51 perspectives for the use of low density genotyped females for GWAS, also for novel phenotypes  
52 that are not currently considered as breeding goals.

53

54 **Key Words:** Lactation curve, Principal Component Analysis, GWAS, LD panel.

55

56

## INTRODUCTION

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

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

68 So far, polygenic **estimated breeding values (EBV)**, daughter yield deviations or deregressed  
69 proofs for traits of interest have been the most frequently used dependent variables in GWAS for

70 dairy traits. The use of genotyped females allows the direct modelling of recorded phenotypes.  
71 Moreover, multiple records are often available for each cow (e.g. multiple measures per lactation,  
72 successive lactations). An appealing, straightforward consequence of the use of repeated records is  
73 that SNP effects can be fitted along the whole lactation curve using longitudinal models (Strucken  
74 et al., 2015).

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

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

96 (Macciotta et al., 2006). Both traits had moderate values of heritability ( $h^2$ ).

97 The aim of the present work was to perform a GWAS on lactation curve traits using model-  
98 free principal component analysis: dependent variables were obtained from the PCA of test-day  
99 records for milk production traits. Cows were genotyped with a low density (7K) SNP panel, as  
100 common in GS programs. All animals belonged to the Italian Simmental, the third-ranked breed in  
101 Italy for milk production (42.133 recorded lactations in 2013 in the Herd book) after Italian  
102 Holstein and Italian Brown (ICAR 2014).

103

104

## MATERIALS AND METHODS

105 Data consisted of 10,605 test-day records for milk production and milk composition traits  
106 (milk yield, fat and protein percentages, and somatic cell score). Records were from 1,515 lactations  
107 of 337 Italian Simmental cows. Animals were farmed in 120 herds. Seven test-day (TD) records  
108 were considered for each lactation. Extra TD for lactations with more than 7 records were deleted.  
109 All cows were genotyped with the 7K Illumina bead-chip. Marker edits were on call rate ( $>.99$ ) and  
110 minor allele frequency ( $>.01$ ). After edits, 6,891 markers were retained for the analysis.

111 Principal component analysis was performed on the seven test day records, separately for each  
112 lactation. Scores of the first two Principal Components (PC1 and PC2) for each lactation were then  
113 calculated by multiplying the row vector of the standardized original variables by the column vector  
114 of the corresponding eigenvector coefficients. Scores were used as dependent variables in the  
115 association study carried out with the following mixed linear model:

$$116 \quad y_{ijklmnop} = \mu + H_i + M_j + Y_k + Par_l + SNP_m + a_n + p_o + e_{ijklmnop}$$

117

118 where  $y$  = PC score;  $\mu$  = overall mean;  $H$  = fixed effect of the  $i$ -th herd,  $M$  = fixed effect of the  $j$ -th  
119 calving month (12 months);  $Y$  = fixed effect of the  $k$ -th calving year (from 2002 to 2013);  $Par$  =  
120 fixed effect of  $l$ -th parity (from 1 to 6,  $>6$ );  $SNP$  = fixed covariable of the  $m$ -th SNP marker  
121 genotype (coded as 0,1,2 according to the copies of the second allele);  $a$  = random additive effect of

122 the n-th animal;  $p$  =random permanent environmental effect of the o-th lactation;  $e$  =random  
123 residual. The animal and permanent environment random effects were assumed to be normally  
124 distributed as  $\mathbf{N}(0, \mathbf{A}\sigma^2_A)$  and  $\mathbf{N}(0, \mathbf{I}\sigma^2_p)$ , respectively, where  $\mathbf{A}$  is the pedigree relationship matrix  
125 and  $\sigma^2_A$  and  $\sigma^2_p$  are additive genetic and permanent environmental variances, respectively. The  
126 mixed model was solved using a REML algorithm implemented in the ASREML software (Gilmour  
127 et al., 2000). Values published by Macciotta et al. (2006) were used as variance priors (PC1=  
128 0.4427, 0.4654, 1.4005; PC2= 0.05264, 0.02954, 0.66050, for  $\sigma^2_A$ ,  $\sigma^2_p$ , and  $\sigma^2_e$ , respectively).

129 Bonferroni corrected significance levels for the SNP effects were calculated to account for  
130 multiple testing: uncorrected  $P$  values were multiplied by the number of tests performed (i.e.,  
131 6,891). SNP were considered significantly associated to the considered trait when the corrected  $P$   
132 value was lower than 0.05.

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

144

145

## RESULTS

146

### *Principal component analysis*



147 The first two principal components extracted from the correlation matrix of the test day  
148 records accounted for most of the original variance. The sum of the first two eigenvalues was 83%  
149 for milk yield, 50% for fat percentage, 57% for protein percentage, and 42% for SCS, respectively  
150 (Table 1). The structure of eigenvectors showed an association between PC1 and all the test-day  
151 records. PC2 was positively associated with the first and negatively with the second part of  
152 lactation, respectively. Such a structure could be observed for all the four considered traits. The first  
153 eigenvalue was always markedly larger than the second in agreement with previous studies  
154 (Macciotta et al., 2006). Differences in magnitude can be observed between traits (Table 1).

155 Figures 1-4 report average lactation patterns for animals grouped according to PC1 and PC2  
156 scores for the four considered traits. In particular, figures 1 (a and b) highlight the role of the two  
157 PC as phenotypic indices of level of production (LEVEL) and lactation curve shape (SHAPE) for  
158 milk yield, respectively. The first two PC have a similar meaning also for fat percentage (Figures 2a  
159 and 2b), protein percentage (Figures 3a and 3b) and somatic cell score (Figures 4a and 4b).

160

### 161 *Association study*

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

167

### 168 *Milk yield*

169 The top significant SNP for milk yield LEVEL (corrected  $P = 0.003$ ) was located on BTA6 at  
170 approximately 89Mb. A strong association between this region and clinical mastitis, milk yield, and  
171 protein percentage was reported in dairy cattle (Sahana et al., 2014). Some interesting genes are  
172 located in the interval of approximately 0.48 Mb calculated around the significant SNP (Table 2).

173 One is the *GC* (Groups specific Component), a gene that encodes for a vitamin D binding protein. It  
174 is involved in several physiological functions as the modulation of inflammatory and immune  
175 response, binding of fatty acids and bone development (Speeckaert et al., 2014). It has been  
176 suggested as a putative candidate for milk yield (Raven et al., 2014) and clinical mastitis (Sahana et  
177 al., 2014). The casein cluster is also located in this genomic region, but it did not fall in the  
178 considered interval. The second SNP in the rank mapped on BTA21, close to the *NRTK3*  
179 (neurotrophic tyrosine kinase, receptor, type 3) locus (Table 3). This gene encodes for a membrane  
180 protein receptor and it is involved in the determinism of a type of breast carcinoma in humans,  
181 (Tognon et al., 2002). Other interesting genes located in this region are two mitochondrial  
182 ribosomal proteins (*MRLP1* and *MRPS11*) that have been reported as candidates for mitochondrial  
183 disorders in humans (Vasta et al., 2009).

184

### 185 ***Fat percentage***

186 The most significant SNP ( $P < 0.001$ ) associated with the fat content LEVEL was located on  
187 BTA23 (Figure 5). A suggestive candidate gene that maps in the window around this marker is the  
188 desmoplakin locus. This gene encodes for a protein involved in the structure of desmosomes,  
189 intercellular junctions that provide tissue integrity at epithelial level (Garrod and Chidgey, 2008).  
190 The second significant marker was located on BTA7 at approximately 71.5 Mb. The clathrin  
191 interactor 1 (*CLINT1*) locus that encodes for a protein which is involved in the vesicle trafficking  
192 (Dodd et al., 2009) maps in this region. This gene has been found to be associated with skin colour  
193 in chicken (Sun et al., 2013).

194 Two significant SNP for LEVEL of fat content were highlighted on BTA14. They both map  
195 in a very dense region where the *DGATI* is located. This gene has a well known major effect on  
196 milk fat in cattle (Grisart et al., 2002). The first marker (ARS-BFGL-NGS-34135) was reported to  
197 be statistically associated to milk yield, fat and protein percentages in a large multibreed study  
198 (Raven et al., 2014). The Zinc Finger Protein 34 (*ZNF 34*) and the Glutamic Pyruvate Transaminase

199 (*GPT*) are also located in the interval around this SNP. Associations between these genes and fat  
200 yield and percentage have been detected in Chinese Holstein (Jiang et al., 2014). The second SNP  
201 (ARS-BFGL-NGS-4939) was highlighted in US (Cole et al. 2011) and German (Wang et al., 2014)  
202 Holsteins, and in a multibreed population (Raven et al., 2014). The window considered for this  
203 marker was fairly large, due to the high linkage disequilibrium of this genomic region (Table 2). It  
204 includes the cytochrome P450, subfamily XI B, polypeptide 1 (*CYP11B1*). This gene has been  
205 suggested as a second relevant QTL on BTA14 affecting fat yield and content in cattle (Jiang et al.,  
206 2014; Kaupe et al., 2007; Mai et al., 2010).

207 A significant SNP for fat percentage mapped on BTA3 at around 99.6 Mb. The glutathione  
208 reductase gene is located in this region. Associations between genes involved in the metabolism of  
209 glutathione and milk yield, fat and protein percentage have been found in cattle (Raven et al., 2014).  
210 Moreover, the reductase gene family has been reported under balancing selection in a recent  
211 comparison between *Bos taurus* and *Bos indicus* genomes (Porto-Neto et al., 2013). The marker  
212 found on BTA 17 (Table 2) defined a 0.53 Mb window where maps the Claudin 5 (*CLDN5*). This  
213 gene encodes for a membrane protein that is a component of the tight junctions. The significant  
214 SNP on BTA2 was found in a region that harbours several interesting genes. One is the Long-chain  
215 Acyl coenzyme A dehydrogenase (*ACAD*) gene, which encodes for a key enzyme of the fatty acid  
216 metabolism in the liver (Lia et al., 2013). Another gene located in this genomic region is the myosin  
217 light chain 1 (*MYL1*), whose expression has been related to the physiological status (lactation vs  
218 puberty) in cattle (Ron et al., 2007).

219 The second significant SNP found on BTA23 was located in a region that harbours some  
220 genes of potential interest. One is the Apolipoprotein B mRNA-editing enzyme catalytic subunit 2  
221 (*APOBEC2*). It is expressed in the muscle and affects muscle fibre ratio and body mass in mice  
222 (Sato et al., 2010). Another is the MyoD family inhibitor (*MDF1*) that was proposed as a suggestive  
223 candidate gene for a QTL that affects fatness traits in pigs (Huang et al., 2011). Finally, also the  
224 interval around the significant marker found on BTA19 included different genes of potential

225 interest. The *ATP2A3* that is involved in calcium mobilization in the cell, could be mentioned. This  
226 gene has been recently reported as a selection signature in Ethiopian cattle populations (Edea et al.,  
227 2014). Another is the olfactory receptor, family 1, subfamily E, member 2-like (*LOC618124*).  
228 Genes of the olfactory receptor family have been found in selection signatures in cattle (Qanbari et  
229 al., 2010).

230

### 231 *Protein percentage*

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

244 The third marker was detected on BTA6. It is placed close to a gene cluster that included the  
245 leucine-rich repeat LGI family, member 2 (*LGI2*), and the DEAH (Asp-Glu-Ala-His) box  
246 polypeptide 15 (*DHX15*), involved in the immune response mechanism. Associations between these  
247 two genes and fat and protein percentages in German Holstein and Fleckvieh have been reported,  
248 respectively (Weikard et al., 2011). Another interesting gene located in this region is the superoxide  
249 dismutase 3 (*SOD3*). It has been found to be associated with intake in beef cattle (Al-Husseini et  
250 al., 2013). The fourth marker affecting the PC1 for protein percentage was located on BTA7, in a

251 region characterized by a relevant LD (Table 2). Several interesting genes could be found in the  
252 calculated window. Of particular interest is the Eukaryotic Translation Elongation Factor 2 (*EEF2*),  
253 involved in milk protein synthesis in the mammary gland as a mediator of the effect of growth  
254 hormone (Hayashi et al., 2009). Moreover, it has been found to be down regulated in cattle  
255 experimentally infected by Bovine tuberculosis (Meade et al., 2007). Another gene of interest in this  
256 region is the integrin beta 1 binding protein 3 (*ITGB1BP3*), that was found under positive selection  
257 in a comparison between Jersey, Guensey, and Zebuine breeds (Porto-Neto et al, 2013). Intervals  
258 surrounding the last two markers for protein percentage, located on BTA13 and BTA5 respectively,  
259 did not include any annotated gene for the bovine genome.

260

### 261 *Somatic cell score*

262 The only significant marker for somatic cell score was located at the beginning of BTA22. In  
263 this region, the vesicular, over expressed in cancer, prosurvival protein 1 (VOPPI) maps. This gene  
264 encodes for a protein that is involved in cellular apoptosis in vertebrates (Pei and Grishin, 2012).

265

266

## DISCUSSION

### 267 *Principal component analysis*

268 PCA performed on test day records for milk production traits was able to synthesise the  
269 main aspects of the lactation pattern, i.e., the general level of production and its shape. The use of  
270 PCA sometimes results in a first component correlated with almost all original variables that  
271 absorbs most of the total variance. In these cases, the eigenvectors of the other extracted PC often  
272 show some variations (values that increase or decrease, or that change sign) in comparison with  
273 PC1 (Stearns et al., 2005). These eigenvector structures allow to infer specific aspects of  
274 relationships between groups of variables that cannot be seen in the first PC (Jombart et al., 2009).  
275 In the case of the lactation curve, the predominance of the first eigenvalue over the second may be  
276 also an expression of the larger heritability of milk yield compared to lactation persistency (Cole

277 and Null, 2009; Cole and VanRaden, 2006). The interpretation of eigenvalues as expression of the  
278 genetic contribution of principal components to phenotypes was also proposed by Kirkpatrick and  
279 Meyer (2004).

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

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

291

## 292 *Association analysis*

293 A relevant number of significant associations were detected in spite of the low density  
294 marker map used. Such a result may be somewhat unexpected because the extent of linkage  
295 disequilibrium between marker and QTL is one of the main factors affecting the power of GWAS  
296 (Powell et al., 2011). On the other hand, a large number of significant markers was reported in a  
297 previous research carried out on females (Cole et al., 2011). Furthermore, some of the most  
298 significant SNP markers found in this study confirmed previous reports on dairy cattle (Cole et al.  
299 2011; Raven et al., 2014; Wang et al., 2012).

300 It is worth remembering that GWAS on males are usually carried out on progeny tested or  
301 on AI bulls, i.e., on the top animals of the breed. A reduction of the genetic variability in  
302 comparison with females could be therefore expected. On the other hand, it has to be remembered

303 that the Italian Simmental is a dual-purpose breed. Bulls are selected by combining performance  
304 test for beef traits and progeny test for dairy traits respectively. Moreover there is limitation in the  
305 number of semen doses that can be used per bull (7,000). These two aspects may have contributed  
306 to reduce the selection pressure and to maintain genetic variation.

307 The kind of analysed traits could be a further cause of results here obtained. Strucken et al.  
308 (2011) found that the use of lactation curve parameters instead of yield data as dependent variables  
309 provided greater power in detecting associations. Authors explained their results with an increase of  
310 the genetic variance of the considered trait. In the present study, the Bonferroni correction of the  
311 SNP significance level was implemented to prevent the occurrence of false positives. Moreover,  
312 population stratification was accounted for by including the polygenic effect in the model.

313 All significant associations were found only for the principal component that described the  
314 level at which the lactation curve is located. No markers were detected for the shape of the lactation  
315 curve. This result is in contrast with previous studies that detected SNP associated to lactation  
316 persistency. Some points could be discussed in order to find possible reasons for these  
317 discrepancies. First of all the kind of dependent variable used. There is a lack of consensus on a  
318 suitable measure of lactation persistency and this trait has been defined in many ways. The measure  
319 of lactation curve shape used in the present paper is completely uncorrelated from milk yield (i.e.  
320  $r_{PC1,PC2}=0$ ), and it explains a reduced quota of variation compared to the LEVEL component. This  
321 may be also an explanation of the absence of common significant markers across traits. On the other  
322 hand, in the paper by Strucken et al. (2011) persistency is defined by using the parameter  $c$  of the  
323 Wilmlink function, which is usually highly correlated with the parameter  $a$ , i.e., the one related with  
324 yield (Macciotta et al., 2005). A second point is represented by the effect of parity. It is widely  
325 acknowledged that first calving cows have a markedly flatter curve in comparison with older  
326 parities. In the paper of Pryce et al. (2010), only primiparous cows were considered. Moreover,  
327 Strucken et al. (2011) found significant associations only when first lactations were analysed  
328 separately from those of older animals. In the present paper, cows of different parity were included

329 in the data set and primiparous represented about 25% of the records. Finally, it should be  
330 remembered that variation of persistency explained by significant SNP in previous researches was  
331 rather low, 1 to 2% in the study of Pryce et al. (2010).

332

### 333 *Gene discovery*

334 The intervals surrounding the significant SNP allowed for the detection of several suggestive  
335 candidate genes. Most of them were found for milk fat percentage. Some of these genes were  
336 already known to affect dairy traits, as the *DGATI*. This is a rather common result for GWAS but it  
337 was somewhat unexpected for Italian Simmental. Pintus et al. (2012) hypothesised that the low  
338 DGV accuracy for fat percentage obtained in Italian Simmental bulls was due to the absence of  
339 allele segregation at the *DGATI* locus in this breed. The frequency of the favourable *DGATI* allele  
340 was >0.99 in a previous work carried out on 95 cows (Scotti et al., 2010). Results of the present  
341 study have been obtained on larger sample. Moreover, the two significant markers found in the  
342 region of BTA14 that harbours the *DGATI* locus have been reported also for other breeds (Cole et  
343 al., 2011; Raven et al., 2011; Wang et al., 2012). A recent study carried out on Italian Simmental  
344 bulls reported a SNP in the promoter region of the *DGATI*, but no associations with dairy traits  
345 were found (Chessa et al. 2015). In any case it is worth remembering that several genes map in this  
346 genomic region. Thus SNP significance could be also due to the effect of other genes.

347 Most of identified putative candidate genes have a biological connection with the lactation.  
348 Examples are those involved in the lipid metabolism (*ACAD*), in protein synthesis (*EEF2*), in the  
349 integrity of the epithelium (*CLDN5* and the desmoplakin), and in the immune response (*DHX15*,  
350 *GC*, and *LCPI*). Of interest is also the gene found on BTA6 that encodes antioxidant enzyme  
351 *SOD3*, whose expression has been found to be downregulated in the mammary tissue of cows fed a  
352 diet rich in polyunsaturated acid (Cortes et al., 2012). This gene is involved in the determinism of  
353 feed intake in cattle (Al-Husseini et al., 2013). Other genes, that map in genomic regions where  
354 selection signatures have been detected, are of more general functions. Examples are olfactory



355 receptors, transmembrane proteins involved cellular processes, and mediators of early development  
356 as *LOC618124*, *VOPPI*, and *ITGB1BP3*, respectively.

357

358

## CONCLUSIONS

359 The GWAS carried out in Italian Simmental cows highlighted some markers statistically  
360 associated with lactation curve traits. In particular, significant associations **were** found for the  
361 principal component that describes the production level at which the lactation curve for different  
362 dairy traits is located. No significant SNP were found for lactation curve shape. **This result**  
363 **disagrees with previous studies on dairy cattle that report significant associations between genomic**  
364 **regions and lactation persistency** Reasons can be found in the kind of measure used, in the dual  
365 purpose aptitude of the Italian Simmental breed, and in the composition of the sample of animals  
366 considered (especially concerning parity). On the other hand, some putative candidate genes  
367 detected in the present study were found to be associated to production traits in previous researches.  
368 In spite of the low density marker map used, a relatively large number of significant markers was  
369 detected. **These results** suggest the use of low density genotyped females for GWAS also for novel  
370 phenotypes (e.g. milk fatty acid spectrum, milk coagulation properties) that are not currently  
371 measured in breeding programs.

372

373

## ACKNOWLEDGEMENTS

374 This research has been funded by the Italian Ministry of Agriculture (Grant INNOVAGEN, **Project**  
375 **Leader** prof. Alessio Valentini).

376

377

378

379

380

381

382 Al-Husseini, W. , C. Gondro, K. Quin, R. M. Herd, J. P. Gibson and Y. Chen. 2013. Expression of  
383 candidate genes for residual feed intake in Angus cattle. *Anim. Genet.* 45: 12–19.

384

385  
386 Barrett, J.C, B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: analysis and visualization of LD  
387 and haplotype maps. *Bioinf.* 21:263-265.  
388  
389 Bolormaa, S., J.E. Pryce , B. J. Hayes , and M. E. Goddard. 2010. Multivariate analysis of a  
390 genome-wide association study in dairy cattle. *J. Dairy Sci.* 93 :3818–3833.  
391  
392 Chessa, S., E.L. Nicolazzi, L. Nicoloso, R. Negrini, R. Marino, D. Vicario, P. Ajmone Marsan, A.  
393 Valentini, B. Stefanon. 2015. Analysis of candidate SNPs affecting milk and functional traits  
394 in the dual-purpose Italian Simmental cattle. *Livest. Sci.*173:1.8.  
395  
396 Cole, J. B., and P. M. VanRaden. 2006. Genetic Evaluation and Best Prediction of Lactation  
397 Persistency. *J. Dairy Sci.* 89:2722–2728.  
398  
399 Cole, J.B., and D.J. Null. 2009. Genetic evaluation of lactation persistency for five breeds of dairy  
400 cattle. *J. Dairy Sci.* 92:2248–2258.  
401  
402 Cole, J.B., G.R. Wiggans, L. Ma, T.S. Sonstegard, T.J. Lawlor Jr., B.A. Crooker, C.P. Van Tassell, J.  
403 Yang, S. Wang, L.K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one  
404 production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows.  
405 *BMC Genom.* 12:408.  
406  
407 Cortes, C., M.F. Palin, N. Gagnon, C. Benchaar, P. Lacasse, and H. V. Petit. 2012. Mammary gene  
408 expression and activity of antioxidant enzymes and concentration of the mammalian lignin  
409 enterolactone in milk and plasma of dairy cows fed flax lignans and infused with flax oil in the  
410 abomasums. *British J. Nutr.* 108: 1390–1398.  
411  
412 Dodd, M. E., J. Hatzold, J. R. Mathias, K. B. Walters, D. A. Bennin, J. Rhodes, J. P. Kanki, A. T.  
413 Look, M. Hammerschmidt, and A. Huttenlocher. 2009. The ENTH domain protein Clint1 is  
414 required for epidermal homeostasis in zebrafish. *Develop.* 136: 2591-2600.  
415  
416 Edea, Z., H. Dadi, S.-W. Kim, J.-H. Park, G.-H. Shin, T. Dessie and K.-S. Kim. 2014. Linkage  
417 disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different  
418 ecological conditions in Ethiopia. *J. Anim. Breed. Genet.* 131: 358–366  
419  
420 Garrod, D., and M. Chidgey. 2008. Desmosome structure, composition and function. *Biochimica et*  
421 *Biophys. Acta* 1778: 572–587.  
422  
423 Gilmour, A. R., B. R. Cullis, S. J. Welhan, and R. Thompson. 2000. *ASREML manual*. New South  
424 Wales Dept. Agric., Orange, Australia.  
425  
426 Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid,  
427 P. Simon, R. Spelman, M. Georges, and R. Snell. (2002) Positional candidate cloning of a QTL in  
428 dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on  
429 milk yield and composition. *Genome Res.* 12: 222-231.  
430  
431 Hayashi, A.A., K. Nones, N. C. Roy, W. C. McNabb, D. S. Mackenzie, D. Pacheco, and S.  
432 McCoard. 2009. Initiation and elongation steps of mRNA translation are involved in the increase in  
433 milk protein yield caused by growth hormone administration during lactation. *J. Dairy Sci.*  
434 92:1889–1899.  
435

- 436 Hayes, B. J., J. Pryce, A.J. Chamberlain, P.J. Bowman, and M.E. Goddard. 2010. Genetic  
437 architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage,  
438 and type in Holstein cattle as contrasting model traits. *Plos Genet.* 6:e1001139.  
439
- 440 Huang, W.H., Z.X. Ma, Z.Y. Xu, Y.Z. Xiong and B. Zuo. 2011. Detection of novel SNPs and  
441 mapping of the fatness QTL on pig chromosome 7q1.1-1.4 region. *Genet. Mol. Res.* 10: 3090-  
442 3097.  
443
- 444 Huang, J., G. Luo, Z. Zhang, X. Wang, Z. Ju, C. Qi, Y. Zhang, C. Wang, R. Li, J. Li, W. Yin, Y. Xu,  
445 S.J. Moisés, J.J. Llorca and J. Zhong. 2014. iTRAQ-proteomics and bioinformatics analyses of  
446 mammary tissue from cows with clinical mastitis due to natural infection with *Staphylococci*  
447 *aureus*. *BMC Genom.* 15:839.  
448
- 449 International Committee for Animal Recording. 2014. Yearly enquiry on the situation of cow milk  
450 recording in ICAR member countries. Results for the years 2012-2013.. ICAR. Rome. 2014.  
451
- 452 Jiang, L., X. Liu, J. Yang, H. Wang, J. Jiang, L. Liu, S. He, X. Ding, J. Liu, and Q. Zhang. 2014.  
453 Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC*  
454 *Genom.* 15:1105.  
455
- 456 Jombart, T., D. Pontier, and A. B. Dufour. 2009. Genetic markers in the playground of multivariate  
457 analysis. *Hered.* 102:330–341  
458
- 459 Kaupe, B., H. Brandt, E. M. Prinzenberg, and G. Erhardt. 2007. Joint analysis of the influence of  
460 CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation,  
461 reproduction, and productive lifespan in German Holstein cattle. *J. Anim. Sci.* 85:11–21.  
462
- 463 Kirkpatrick, M., and K. Meyer. 2004. Direct estimation of genetic principal components: simplified  
464 analysis of complex phenotypes. *Genet.* 168:2295-2004.  
465
- 466 Lemay, D.G., D.J. Lynn, W.F. Martin, M.C. Neville, T. M. Casey, G. Rincon, E.V. Kriventseva, W.  
467 C. Barris, A.S. Hinrichs, A.J. Molenaar, K.S. Pollard, N.J. Maqbool, K. Singh, R. Murney, E.M.  
468 Zdobnov, R.L. Tellam, J.F. Medrano, J. B. German, and M. Rijnkels. 2009. The bovine lactation  
469 genome: insights into the evolution of mammalian milk. *Genome Biology* 2009, 10:R43  
470
- 471 Lia, P., Y. Liu, Y. Zhang, M. Long, Y. Guo, Z. Wang, X. Li, C. Zhang, X. Li, J. He, G. Liu. 2013.  
472 Effect of Non-Esterified Fatty Acids on Fatty Acid Metabolism-Related Genes in Calf Hepatocytes  
473 Cultured in Vitro. *Cell. Physiol. Biochem.* 32:1509-1516  
474  
475
- 476 Lourenco, D.A.L., I. Misztal, S. Tsuruta, I. Aguilar, E. Ezra, M. Ron, A. Shirak, and J. I.  
477 Weller. 2014. Methods for genomic evaluation of a relatively small genotyped dairy population and  
478 effect of genotyped cow information in multiparity analyses. *J. Dairy Sci.* 97 :1742–1752.  
479
- 480 Mai, M. D., G. Sahana, F. B. Christiansen, and B. Guldbandsen. 2010. A genome-wide association  
481 study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism  
482 chip. *J. Anim. Sci.* 2010. 88:3522–3528.  
483
- 484 Macciotta, N. P. P. , D. Vicario, and A. Cappio-Borlino. 2005. Detection of Different Shapes of  
485 Lactation Curve for Milk Yield in Dairy Cattle by Empirical Mathematical Models. *J. Dairy Sci.*  
486 88:1178–1191

487  
488 Macciotta, N.P.P., D.Vicario, and A. Cappio-Borlino. 2006. Use of multivariate analysis to extract  
489 latent variables related to level of production and lactation persistency in dairy cattle. *J. Dairy Sci.*  
490 89: 3188-3194.  
491  
492 Macciotta, N.P.P. , C.Dimauro, S.P.G. Rasso, R. Steri, G. Pulina. 2011. The mathematical  
493 description of lactation curves in dairy cattle. *It. J. of Animal Sci.* 11:213-223.  
494  
495 Meade K.G., E. Gormley, M.B. Doyle, T. Fitzsimons, C. O'Farrelly, E. Costello, J. Keane, Y. Zhao,  
496 and D.E. MacHugh. 2007. Innate gene repression associated with *Mycobacterium bovis* infection in  
497 cattle: toward a gene signature of disease. *BMC Genom.* 8:40  
498  
499 Mohan, S., Y. Hu, B. Edderkaoui. 2013. Identification of Gender-Specific Candidate Genes that  
500 Influence Bone Microarchitecture in Chromosome 1. *Calc. Tiss. Int.* 92: 362-371  
501  
502 Olori, V.E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1999. Fit of standard models of the  
503 lactation curve to weekly records of milk production of cows in a single herd. *Livest. Prod. Sci.* 58:  
504 55-63.  
505  
506 Pei, J., and N. V. Grishin. 2012. Unexpected diversity in Shisa-like proteins suggests the importance  
507 of their roles as transmembrane adaptors. *Cell. Signall.* 24:758–769  
508  
509 Pintus, M. A., G. Gaspa, E. L. Nicolazzi, D. Vicario, A. Rossoni, P. Ajmone-Marsan, A. Nardone, C.  
510 Dimauro, and N.P.P. Macciotta. 2012. Prediction of genomic breeding values for dairy traits in  
511 Italian Brown and Simmental bulls using a principal component approach. *J. Dairy Sci.* 95:3390–  
512 3400  
513  
514 Porto-Neto, L.R., T.S. Sonstegard, G.E. Liu, D.M. Bickhart, M.V.B. Da Silva, M.A. Machado, Y.T.  
515 Utsunomiya, J.F. Garcia, C. Gondro, and C.P. Van Tassell. 2013. Genomic divergence of zebu and  
516 taurine cattle identified through high-density SNP genotyping. *BMC Genom.* 14:876  
517  
518 Powell, J. E., A. Kranis, J. Floyd, J. C. M. Dekkers, S. Knott, and C. S. Haley. 2011. Optimal use of  
519 regression models in genome-wide association studies. *Anim. Genet.* 43: 133–143.  
520  
521 Pryce, J.E., M. Haile-Mariam, K. Verbyla, P.J. Bowman, M.E. Goddard, and B.J. Hayes. 2010.  
522 Genetic markers for lactation persistency in primiparous Australian dairy cows *J. Dairy Sci.* 93:  
523 2202-2214  
524  
525 Qanbari, S., E.C.G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi and H. Simianer. 2010.  
526 A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41: 377–389  
527  
528 Raven, L.A. , B.G. Cocks, and B.J Hayes. 2014. Multibreed genome wide association can improve  
529 precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genom.*  
530 15:62.  
531  
532 Ron, M., G. Israeli, E. Seroussi, J. I Weller, J.P. Gregg, M. Shani, and J. F. Medrano. 2007.  
533 Combining mouse mammary gland gene expression and comparative mapping for the identification  
534 of candidate genes for QTL of milk production traits in cattle. *BMC Genom.* 8:183  
535  
536 Sahana, G., B. Gulbrandsen, B. Thomsen, L-E. Holm, F. Panitz, R.F. Brøndum, C. Bendixen,  
537 M.S. Lund. 2014. Genome-wide association study using high-density single nucleotide

538 polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *J.*  
539 *Dairy Sci.* 97: 7258–7275  
540

541 Sato, Y., H.C. Probst, R. Tatsumi, Y. Ikeuchi, M.S. Neuberger, and C. Rada. 2010. Deficiency in  
542 APOBEC2 Leads to a Shift in Muscle Fiber Type, Diminished Body Mass, and Myopathy. *J. Biol.*  
543 *Chem.* 285:7111–7118.  
544

545 Scotti, E. , L. Fontanesi, F. Schiavini, V. La Mattina, A. Bagnato, V. Russo. 2010. Short  
546 communication: DGAT1 p.K232A polymorphism in dairy and dual purpose Italian cattle breeds. *It.*  
547 *J. Anim. Sci.* 9:e16  
548

549 Shin, D.H., H.J. Lee, S. Cho, H.J. Kim, J.Y. Hwang, C.K. Lee, J.Y. Jeong, D. Yoon, and H. Kim.  
550 2014. Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at  
551 the population level. *BMC Genom.* 15:240  
552

553 Speeckaert, M.M., R. Speeckaert, N. van Geel, J.R. Delanghe. 2014. Chapter One – Vitamin D  
554 Binding Protein: A Multifunctional Protein of Clinical Importance. *Adv. Clin. Chem.* 63: 1-57.  
555  
556

557 Stearns, T.M., J.E. Beever, B.R. Southey, M. Ellis, F.M. McKeith, and S.L. Rodrigues-Szas. 2005.  
558 Evaluation of approaches to detect quantitative loci for growth, carcass and meat quality on swine  
559 chromosomes 2, 6, 13, and 18. II. Multivariate and principal component analyses. *J. Anim. Sci.*  
560 83:2471-2481.  
561

562 Strucken, E.M., D.J. de Koning, S.A. Rahmatalla, and G.A. Brockmann. 2011. Lactation curve  
563 models for estimating gene effects over a timeline. *J. Dairy Sci.* 94:442–449  
564 Strucken, E.M., R.H. Bortfeldt, D.J. de Koning, and G.A. Brockmann. 2012. Genome-wide  
565 associations for investigating time-dependent genetic effects for milk production traits in dairy  
566 cattle. *Anim. Genet.* 43:375–382.  
567

568 Strucken, EM, Y.C.S.M. Laurenson, and G.A. Brockmann. 2015. Go with the flow—biology and  
569 genetics of the lactation cycle. *Front. Genet.* 6:118. doi: 10.3389/fgene.2015.00118  
570

571 Sun, Y., G. Zhao, R. Liu, M. Zheng, Y. Hu, D Wu, L. Zhang, P. Li and J. Wen. 2013. The  
572 identification of 14 new genes for meat quality traits in chicken using a genome-wide association  
573 study. *BMC Genom.*14:458  
574

575 Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex  
576 traits as variables over time. *Anim. Genet.* 45:322–328.  
577

578 Tognon, C., S.R. Knezevich, D. Huntsman, C.D. Roskelley, N. Melnyk, J.A. Mathers, L. Becker, F.  
579 Carneiro, N. MacPherson, D. Horsman, C. Poremba, P. H.B Sorensen. 2002. Expression of the  
580 ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell*  
581 2:367–76.  
582

583 Thomasen, J.R., A.C. Sørensen, M. S. Lund, and B. Guldbandsen. 2014. Adding cows to the  
584 reference population makes a small dairy population competitive. *J. Dairy Sci.* 97 :5822–5832.  
585

586 Vasta, V., S.B. Ng, E.H. Turner, J. Shendure, and S. Houn Hahn. 2009. Next generation sequence  
587 analysis for mitochondrial disorders. *Genom. Med.* 1:100.  
588

589 Zimin, A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.  
590 P. Van Tassell, T.S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S.L.  
591 Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos Taurus*. *Genom. Biol.* 10:R42  
592  
593 Wang, X., C. Wurmser, H. Pausch, S. Jung, F. Reinhardt, J. Tetens, G. Thaller, R. Fries.2104.  
594 Identification and Dissection of Four Major QTL Affecting Milk Fat Content in the German  
595 Holstein-Friesian Population. *PlosOne* 7:e40711  
596  
597 Weikard, R., P. Widmann, J. Buitkamp, R. Emmerling, and C. Kuehn. 2011. Revisiting the  
598 quantitative trait loci for milk production traits on BTA6. *Anim. Genet.* 43: 318–323  
599  
600 Wiggans, G.R., T.A. Cooper, P.M. VanRaden, and J.B. Cole. 2011. Technical note: Adjustment of  
601 traditional cow evaluations to improve accuracy of genomic predictions. *J. Dairy Sci.* 94 :6188–  
602 6193  
603  
604 Wiggans, G.R., T.A. Cooper, C.P. Van Tassell, T.S. Sonstegard, and E.B. Simpson. 2013. Technical  
605 note: Characteristics and use of the Illumina BovineLD and GeneSeek Genomic Profiler low-  
606 density bead chips for genomic evaluation. *J. Dairy Sci.* 96 :1258–1263.  
607  
608  
609  
610  
611

612 Table 1. Eigenvectors and eigenvalues of the first two principal components (PC1 and PC2)  
 613 extracted for the seven test day records for milk yield, fat and protein percentage, somatic cell  
 614 score.

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

615  
 616  
 617

618  
619  
620

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

Marker	BTA	Position (bp)	Trait <sup>1</sup>	P	Interval
BTB-01654826	6	88,891,318	PC1 MY	0.003	±242,276
Hapmap38505-BTA-51760	21	19,193,100	PC1 MY	0.015	±690,747
ARS-BFGL-NGS-11659	23	47,677,534	PC1 FP	0.0003	±5,663
ARS-BFGL-NGS-75852	7	71,545,383	PC1 FP	0.001	± 214,702
ARS-BFGL-NGS-34135	14	1,675,278	PC1 FP	0.003	± 783,698
ARS-BFGL-NGS-18926	3	99,592,696	PC1 FP	0.009	± 2,160,054
ARS-BFGL-NGS-24012	17	74,948,921	PC1 FP	0.016	± 267,855
ARS-BFGL-NGS-4939	14	1,801,116	PC1 FP	0.029	± 2,574,789
Hapmap34329-BES11_Contig247_1378	2	98,446,391	PC1 FP	0.031	± 132,872
Hapmap41022-BTA-55560	23	15,147,471	PC1 FP	0.044	± 1,729,322
Hapmap58587-ss4652997	19	24,972,085	PC1 FP	0.05	± 256,339
BTB-01225907	16	32,653,232	PC1 PP	0.0205	± 10,322,590
ARS-BFGL-NGS-55674	12	16,069,827	PC1 PP	0.0215	± 373,760
ARS-BFGL-NGS-97136	6	45,909,053	PC1 PP	0.0240	± 868,964
UA-IFASA-8256	7	21,704,630	PC1 PP	0.0341	± 1,619,254
BTA-102818-no-rs	13	83,632,355	PC1 PP	0.0423	± 244,086
Hapmap50090-BTA-75536	5	11,284,482	PC1 PP	0.0492	± 28,293
ARS-BFGL-NGS-84222	22	531,301	PC1 SCS	0.0360	± 43,004

621  
622  
623  
624  
625  
626  
627

<sup>1</sup>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 somatic cell score test data)



628  
629  
630

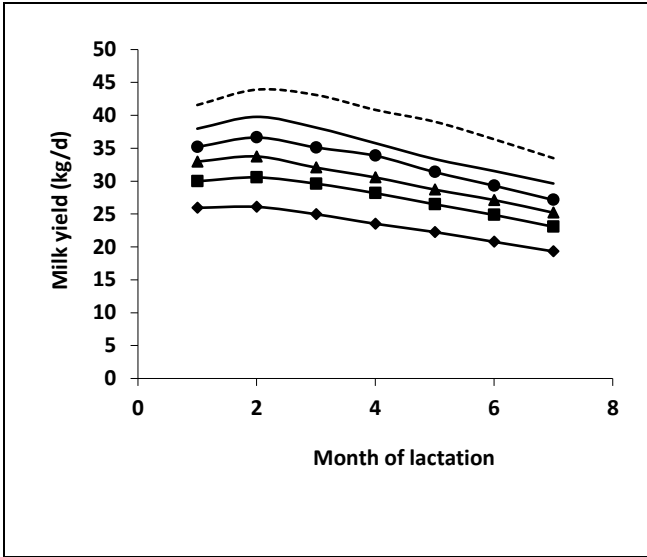


Figure 1a.

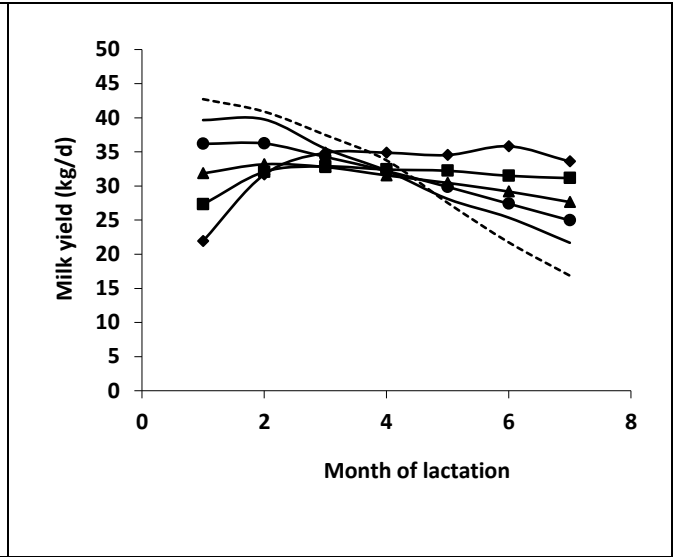


Figure 1b.

631

632

633

634

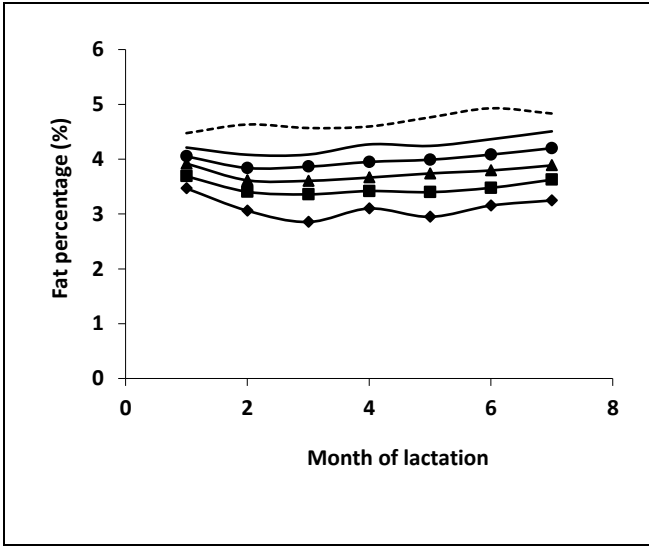


Figure 2a.

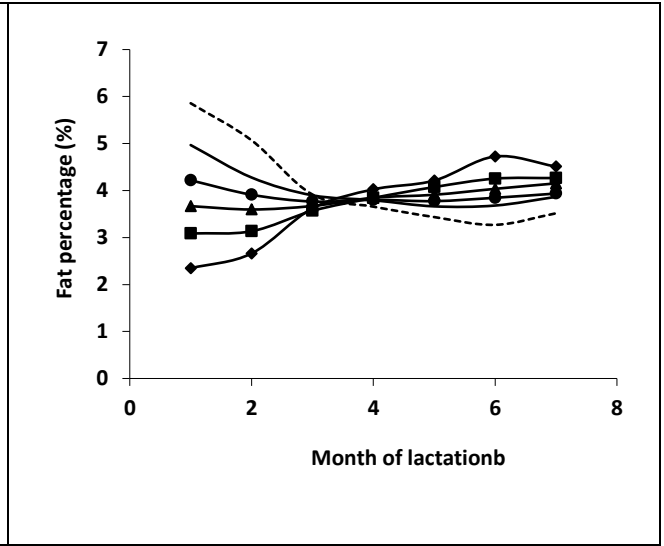


Figure 2b.

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653  
654

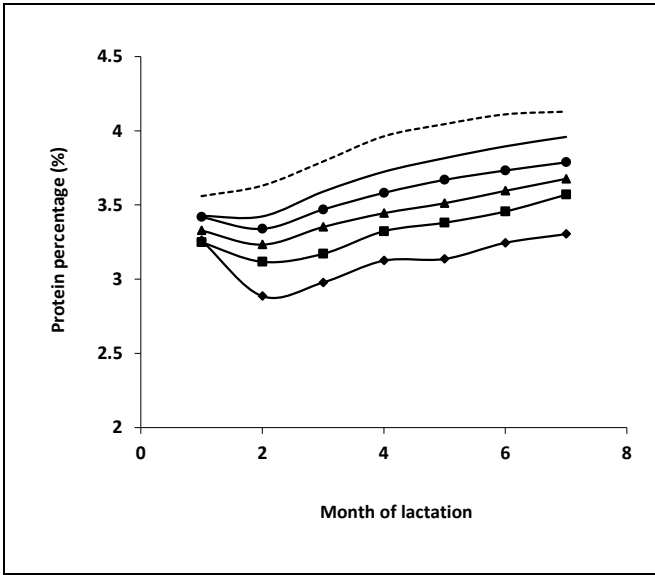


Figure 3a

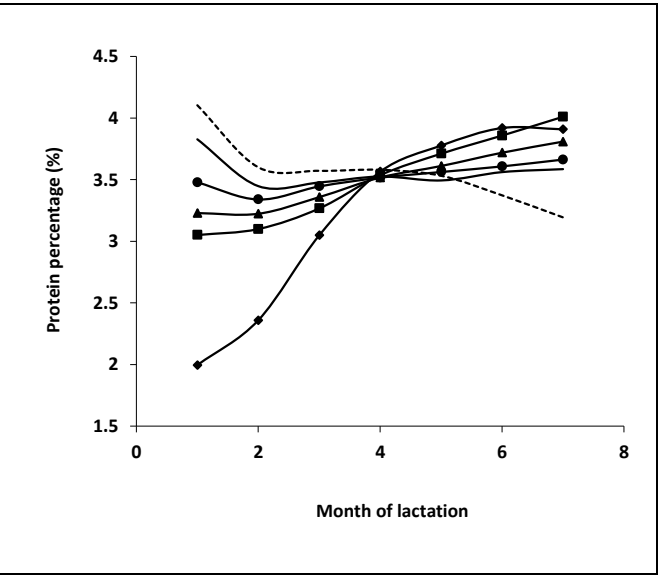


Figure 3b

655  
656  
657  
658  
659  
660

661  
662  
663

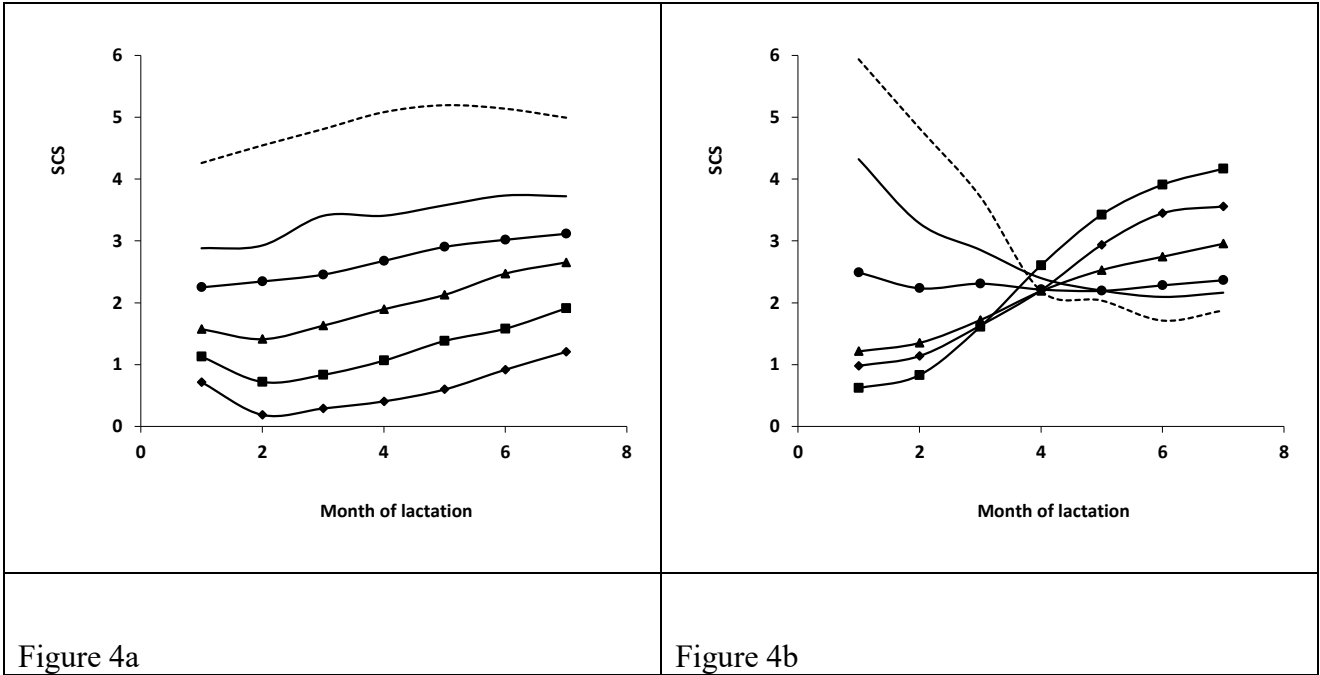


Figure 4a

Figure 4b

664  
665  
666  
667  
668

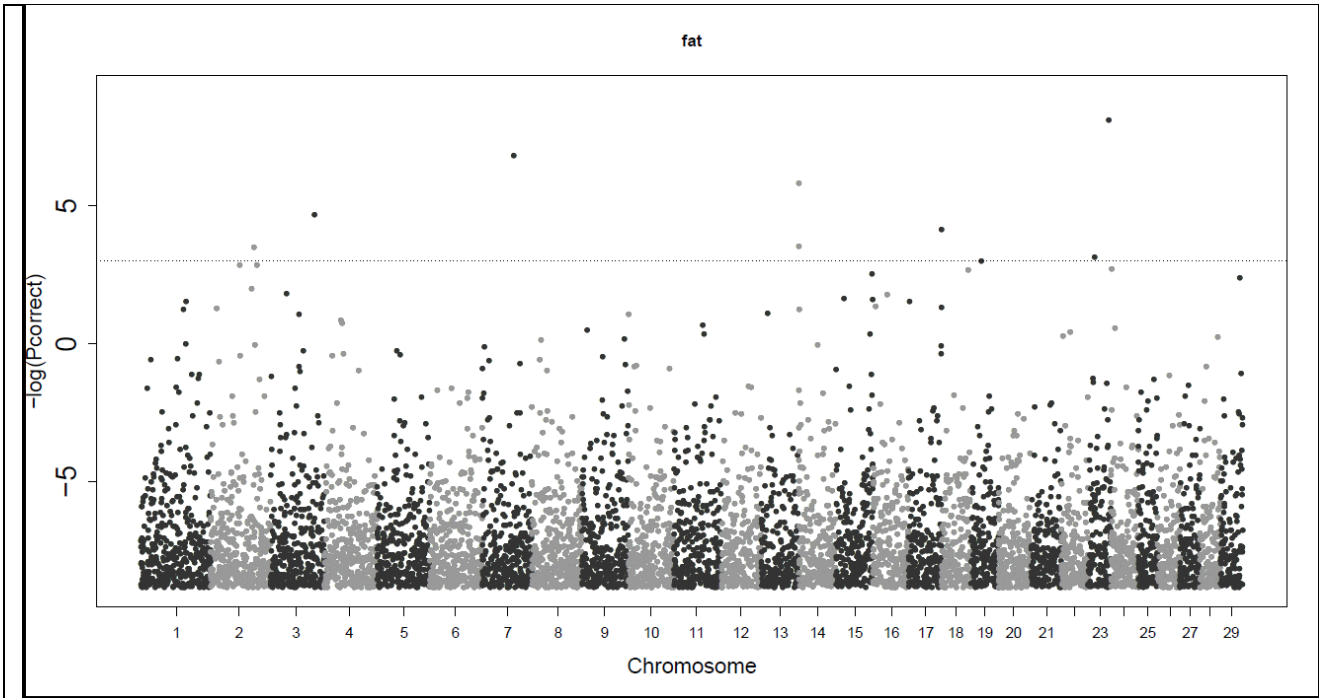


Figure 5

669  
670  
671  
672  
673  
674  
675  
676  
677  
678

679

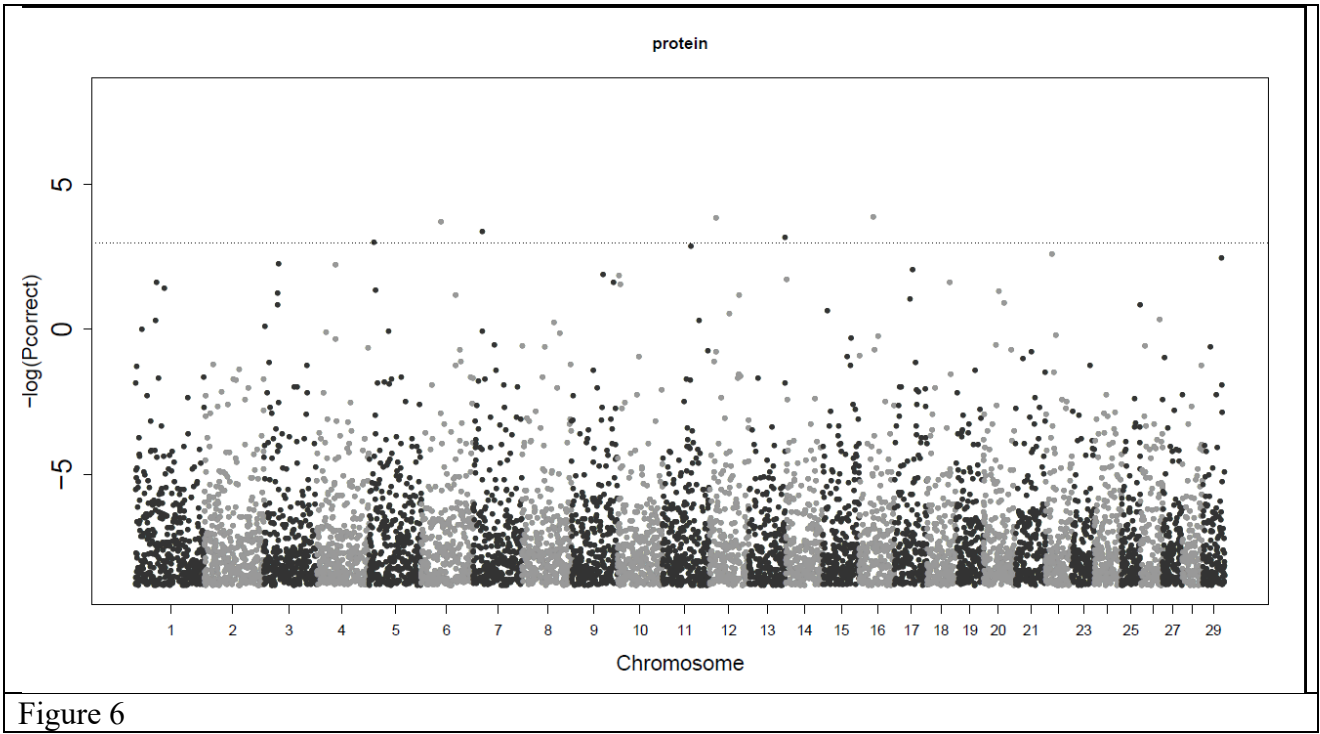


Figure 6

680

681

682

683

684

685

686

687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732

## Captions of figures

Figure 1a. Average lactation curves for milk yield of groups of cows of different PC1 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.

Figure 1b. Average lactation curves for milk yield of groups of cows of different PC2 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.

Figure 2a. Average lactation curves for fat percentage of groups of cows of different PC1 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.

Figure 2b. Average lactation curves for fat percentage of groups of cows of different PC2 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.

Figure 3a. Average lactation curves for protein percentage of groups of cows of different PC1 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.

Figure 3b. Average lactation curves for protein percentage of groups of cows of different PC2 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.

Figure 4a. Average lactation curves for somatic cell score of groups of cows of different PC1 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.

Figure 4b. Average lactation curves for somatic cell score of groups of cows of different PC2 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.

Figure 5. Genome-wide association study of the scores of the first principal component (LEVEL) for fat percentage. The dashed line corresponds to a Bonferroni corrected significance level of 0.05

Figure 5. Genome-wide association study of the scores of the first principal component (LEVEL) for protein percentage. The dashed line corresponds to a Bonferroni corrected significance level of 0.05