GENETICS OF A PCA BASED HEAT TOLERANCE MEASURE

Interpretive summary

Derivation and GWAS of a principal component–based measure of heat tolerance in dairy cattle

by Macciotta et al.

Principal component analysis was used to derive new measures of heat tolerance from milk yield and composition in dairy cattle. Two new variables were extracted: They exhibited genetic variability and were associated to genomic regions that harbor some interesting candidate genes Results suggest their use as a indexes of thermotolerance in dairy cattle breeding schemes.

12 **GENETICS OF A PCA BASED HEAT TOLERANCE MEASURE**

13

14 **Derivation and GWAS of a principal component – based measure of heat tolerance in dairy**

15 **cattle**

16

17 N. P. P. Macciotta[*1], S. Biffani[†], U. Bernabucci[‡], N. Lacetera[‡], A. Vitali[‡], P. Ajmone-Marsan[¥] , and

18 A. Nardone[‡1]

19

20 [*] Dipartimento di Agraria,Università di Sassari, Sassari, Italy

21 [†]Associazione Italiana Allevatori, Roma, Italy

22 [‡]Dipartimento di Scienze Agrarie e Forestali, Università degli Studi della Tuscia-Viterbo, Italy

23 [¥]Istituto di Zootecnica, Università Cattolica del Sacro Cuore, Piacenza, Italy

24

25

26 [1]Corresponding authors: Nicolò P. P. Macciotta, email: macciott@uniss.it; Alessandro Nardone, email:

27 nardone@unitus.it

28

29

ABSTRACT

Heat stress represents a key factor that negatively affects the productive and reproductive performance of farmed animals. In the present work, a new measure of tolerance to heat stress for dairy cattle was developed using Principal Component Analysis. Data were 590,174 test day records for milk yield, fat and protein percentages, and somatic cell score of 39,261 Italian Holstein cows. Test day (TD) records adjusted for main systematic factors were grouped into eleven temperature humidity index (THI) classes. Daughter trait deviations (DTD) were calculated for 1,540 bulls as means of the adjusted TD records for each THI class. Principal Component Analysis was performed on the DTD for each bull. The first two principal components (PC) explained from 42 to 51% of the total variance of the system across the four traits. The first PC, named LEVEL was interpreted as a measure of the level at which the DTD curve was located. The second, named SLOPE, synthesized the behavior of the DTD pattern. Heritability of the two component scores was moderate to high for LEVEL across all traits (range 0.23- 0.82) and low to moderate for SLOPE (range 0.16-0.28). For each trait, phenotypic and genetic correlations between LEVEL and SLOPE were equal to zero. A genome-wide association analysis was carried out on a sub-sample of 423 bulls genotyped with the Illumina 50K bovine bead chip. Two single nucleotide polymorphisms (SNP) were significantly associated to SLOPE for milk yield, four to LEVEL for fat percentage, and two for protein percentage, respectively. The gene discovery carried out considering windows of 0.5 Mb surrounding the significant markers highlighted some interesting candidate genes. Some of them have been already associated to the mechanism of heat tolerance as the heat shock transcription factor (*HSF1*) and the malonyl-CoA-acyl carrier protein transacylase (*MCAT*). The two PC were able to describe the overall level and the slope of response of milk production traits across increasing levels of THI index. Moreover, they exhibited genetic variability and were genetically uncorrelated. These features suggest their use as measures of thermo tolerance in dairy cattle breeding schemes.

Keywords: Heat tolerance, Principal Component Analysis, heritability, GWAS, Dairy cattle

INTRODUCTION

The improvement of animal ability to cope with adverse environmental conditions is one of the great challenges of animal breeding for the future (Bernabucci et al., 2010). Among the traits that contribute to define animal adaptability to environmental variation, tolerance to heat stress plays a major role. Heat stress can be defined as the condition where the animal is not able to adequately dissipate the excess of endogenous or exogenous heat to maintain the body thermal balance (Bernabucci et al., 2014). In dairy cattle it is known that heat stress results in relevant economic losses due to reduced milk production and reproduction performances (Nardone et al., 2010; Aguilar et al., 2010; Biffani et al. 2016). The increasing concern about tolerance to heat stress for dairy animals also in temperate areas is a consequence of both climate changes and higher metabolic heat production by high yielding animals (Kadzere et al. 2002; Hansen, 2007; Segnalini et al., 2011).

If tolerance to heat stress is a quite straightforward concept, its systematic measure remains problematic. On the other hand a quantification of this trait is fundamental if it has to be considered a potential selection goal in breeding programs.

Some physiological traits are related to the ability of the animal to cope with heat stress. For example, rectal temperature and respiration rate increase when animals are exposed to warm environment (Garner et al., 2016; Perano et al., 2015; Dikmen et al., 2012).These traits exhibit a genetic component: for example, a moderate heritability and associations with single nucleotide polymorphisms (**SNP**) and candidate genes have been reported for rectal temperature (Dikmen et al., 2013; 2015). However, the inclusion of these heat tolerance indicator traits in large-scale phenotype recording systems for selecting thermo-tolerant animals appears rather problematic in terms of logistics and costs. An alternative is to evaluate heat tolerance by measuring changes of milk production traits under warm environmental conditions (Carabano et al., 2016; Hammami et

al., 2015; Nguyen et al., 2016). In dairy cattle populations involved in selection programs, milk production data could be easily retrieved from dairy recording systems and associated to climate data provided by weather stations. The variable most frequently used to evaluate heat stress conditions is the temperature-humidity index (**THI**). The approach commonly used to evaluate heat tolerance relies on the so-called broken line model (Misztal, 1999). It assumes the existence of a comfort zone limited by an upper threshold value (**TH_0**) beyond which the production linearly decreases as THI increases (Bernabucci et al., 2014; Carabano et al., 2014).

In statistical models tolerance to heat stress might be fitted according to a reaction norm model (Kolmodin and Bijma, 2004), where the phenotype is expressed as a linear function of an environmental variable (for example THI or temperature). Very often, the environmental variable effect is a dummy variable, set to zero when $THI<TH_0$ and to $THI-TH_0$ when $THI>TH_0$ (Bernabucci et al., 2014). Some studies adopted the fixed value of 72 for TH_0 (Aguilar et al., 2009; Bohmanova et al., 2008; Ravagnolo and Misztal, 2000) but recently different TH_0 have been estimated across traits, parities and geographical regions (Bernabucci et al., 2014; Biffani et al., 2016).

Some studies on tolerance of heat stress have used individual production curves along different THI levels corrected for fixed factors as a measure of heat tolerance (Carabano et al., 2016; Hayes et al., 2009). Average curves of bull progeny for milk production traits across different THI levels, named as daughter trait deviations (**DTD**), have been recently used as phenotypes in a genomic selection study on tolerance to heat stress (Nguyen et al., 2016).

For genetic purposes, individual effects for heat tolerance are usually fitted with an intercept and a slope, representing the overall level of production and the response of the animal to heat stress, respectively. Main concerns about these approaches are on the use of a common threshold across all animals and the assumption of linearity for the production decay after TH_0 (Bernabucci et al., 2014; Carabano et al., 2014). On the other hand, estimation of individual thresholds (Sanchez et al., 2009) is more realistic even though more computationally demanding. Individual change points

of production patterns for increasing THI levels have been fitted also with Legendre polynomials in Random Regression Models (Brugemann et al., 2012; Carabano et al, 2014; 2016)

Several papers that evaluated the effect of heat stress on milk reported an unfavorable genetic relationship between production and heat tolerance (Hammami et al., 2015; Sanchez et al., 2009; Bernabucci et al., 2014). These results were confirmed also by the strong negative correlations (-0.85 and -0-75) between genomic breeding value for milk DTD derived heat tolerance and EBV for milk yield in Australian Holsteins and Jerseys respectively (Nguyen et al., 2016). Such correlation is the result of the increased metabolic heat production that occurs in high producing cows and that exacerbate the effects of the external heat. This represents a severe constraint to an efficient selection for improving heat tolerance without negative consequences on production. The aggregation of the two traits into a selection index may help selection, even though the definition of optimal economic weights could remain theoretical issue and the negative correlation undoubtedly will reduce the selection response on each individual trait. An alternative could be the use of a measure of tolerance to heat stress that is not correlated with production levels. The use of a model free-approach, able to disentangle main features of DTD without imposing specific constraints is an appealing option for assessing proper variables to study tolerance to heat stress. Principal component analysis (**PCA**) is a multivariate statistical technique able to synthesize complex patterns as the lactation curves for dairy traits in two variables with a clear technical meaning (Macciotta et al., 2006; 2015). PCA can be therefore conveniently used to analyze DTD curves for extracting new variables able to synthesize the pattern.

In the present work a PCA approach was tested to derive indicator variables of tolerance to heat stress from milk production data in dairy cattle. Moreover, a genome wide association study (**GWAS**) using MD (medium) density (50K) SNP panel was used for investigating the genetic determinism of these new variables.

MATERIALS AND METHODS

133 ***Data***

134 Data were 590,174 test day (**TD**) records for milk yield (**MY**), fat (**FP**) and protein (**PP**)
135 percentages, and somatic cell score (**SCS**) (Ali and Shook, 1980) of 39,261 Italian Holstein cows
136 (first, second, and third parity) from 484 farms, collected from 2001 to 2007. Data were recorded by
137 the Italian Breeders Association (**AIA**) according to International Committee for Animal Recording
138 (**ICAR**) standards (http://www.icar.org/Documents/Rules%20and%20regulations/
139 Guidelines/Guidelines_2011.pdf). Age at calving classes were established for each parity according
140 to the following thresholds: 20 to 36 (17 classes), 31 to 50 (20 classes), and 42 to 65 (24 classes)
141 month of age for first-, second-, and third-parity cows, respectively. All cows had first lactation data
142 and a minimum of 8 Test Day (**TD**) records per lactation (from 5 to 305 DIM). A minimum of 24
143 records per herd-year of calving were required. Cows were sired by 4,184 AI bulls.

144 Daily weather information were collected from 35 meteorological stations located no more
145 than 5 kilometers from the considered herd. THI index (Kelly and Bond, 1971) was then calculated
146 as:

147 $THI = (1.8 \times AT + 32) - (0.55 - 0.55 \times RH) \times [(1.8 \times AT + 32) - 58)]$

148 where AT is the maximum daily temperature expressed in Celsius degrees and RH is the minimum
149 relative humidity expressed in percentage.

150

151 ***Statistical analysis***

152 TD records were firstly analyzed with the following mixed linear model

153 $$y = month(year) + age + dim*parity + herd(year) + e \quad [1]$$

154 where

155 y = is the record for MY, FP, PP, or SCS

156 month(year) = is the fixed effect of the month of calving (12 months) nested within the year of
157 calving (7 years 2001-2007);

158 age = is the fixed effect of age class in months (61 classes, from 20 to 65 months);

dim*parity = is the interaction between the fixed effect of the days in milk class (10 intervals of 30 days each) and the fixed effect of parity (3 parities: 1-3)

herd(year) = is the random effect of the herd (458) nested within calving year

e = is the random residual

Residuals of model [1] are therefore production data adjusted for main systematic factors except from additive genetic and THI effects. On the basis of THI values, records were grouped into eleven THI classes (1=50-52, 2=53-54,….11=>79). Distribution of records across THI classes is reported in figure 1. Means of residuals were calculated for each bull and THI class for obtaining Daughter Trait Deviations (DTD) (Nguyen et al., 2016). DTD plotted against the THI class express the sensitivity of bull's daughter production performance for increasing THI levels.

The following step was to derive a measure able to summarize the shape of these curves that could be used as dependent variable in a GWAS. PCA was carried out on the eleven points of the DTD curves, considered as different variables: for example the first bull for milk yield had eleven records (i.e., DTD1_MY, DTD2_MY,…DTD11_MY). Only bulls that had the complete set of eleven DTD values for each trait were considered (1,540 for MY, 1,513 for FP, 1,536 for PP and 1,535 for SCS respectively). PCA was carried out using the SAS PRINCOMP procedure (SAS, 2008). The number of PC to be retained was based on their eigenvalue, and on their relationships with the original variables.

Principal component (**PC**) scores were then calculated for each bull and treated as new phenotypes for performing either genetic parameter estimation and genome wide association analysis (GWAS).Variance components for PC scores were estimated with the following multi-trait animal model:

$$y = \mu + animal + e$$

where **y is** a vector of PC scores for MY, FP, PP, and SCS, respectively, μ is the overall mean; *animal* is the random additive genetic effect; and **e** is the residual term. The following (co)variance structure was assumed for the random effects:

$$Var\begin{vmatrix} a \\ e \end{vmatrix} = \begin{vmatrix} A \otimes G_0 & 0 \\ 0 & I \otimes R_0 \end{vmatrix}$$

where **A** is the numerator relationship matrix; G_0 the matrix of (co)variances for additive effects and $\mathbf{R_0}$ is a diagonal matrix of residual variances corresponding to each trait. The pedigree file had 21,685 animals, including the 1,540 sires with DTDs in the data set.

The model was solved using the program AIREML90 (Misztal et al., 2002). Considering that PC scores were calculated starting from average yields per bull, the ratio

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2}$$

where σ^2_A and σ^2_E are the additive genetic and the residual variances respectively. represents an approximation of the true heritability because averaging affects the variability of the response (different number of TD records per bull). Thus obtained values have been properly called as pseudo-heritability.

Of the 1,540 bulls considered for the PC score calculation, 423 were genotyped with the Illumina 50K bovine bead chip. Monomorphic SNPs (7,140) and SNPs with a call-rate < 95% (1,045) were discarded. In total, 45,546 SNPs were retained for the analysis. Genome-wide scan was performed on PCA scores with the GenABEL R package, using the GRAMMAR procedure. First, an additive polygenic model was fitted to obtain individual residuals using the genomic relationship matrix. Then, SNP association was tested using a linear model on residuals of the first step. The SNP statistical significance was corrected for the stratification of the population using the Genomic Control (**GC**) option (Amin et al., 2007). The GC corrected P values (GC_P_i) were further corrected for multiple testing using either i) the Bonferroni correction, obtained as GC_P_i*m (where

m is the number of performed tests); ii) and calculating the False Discovery Rate (**FDR**), as (GC_P_i*k)/ m_0, where m_0 is the number of tests having the GC P values lower or equal to GC_P_i.A marker was declared significantly associated to a trait when the FDR was <0.10.

Gene discovery analysis was carried out considering windows of 0.5 Mb surrounding the significant marker (0.25Mb up and down stream respectively). Genes were derived from UCSC Genome Browser Gateway (http://genome.ucsc.edu/). SNP and gene positions were obtained from the UMD3.1 bovine genome assembly (Zimin et al., 2009).

RESULTS

Principal Component Analysis

Eigenvectors and eigenvalues of the first two PC extracted from all the four considered phenotypes are reported in Table 1. The first two PC explained from 42 to 51% of the total variance of the system across the four traits. The choice of retaining only the first two PC was motivated by the magnitude of single eigenvalues, even though the amount of explained variance was not particularly relevant. A common criterion used for retaining PC is that the eigenvalue should be greater than one. In the present work, for all the four traits only the first eigenvalue fulfilled this requirement and the second was very close to one (Table 1).

The first principal component (**PC1**) showed positive and moderate eigenvector coefficients or loadings (ranging from 0.25 to 0.35) with all the original variables (Table 1). Thus PC1 can be considered as a measure of the level at which the curve is located and it was named LEVEL. Bulls with large scores for this component have their DTD pattern located on a higher level. The second principal component (**PC2**) exhibited larger loadings (up to 0.76 for DTD_FP) with both positive and negative signs. In particular, PC2 showed positive values for the first part of THI interval and negative for the second part for MY, SCS and, even if less definite, for FP, respectively. On the contrary, PC2 for PP showed the opposite trend. For its structure, the PC2 defined the shape of the DTD curve. Therefore larger values (or smaller in the case of protein content) of PC2 scores

characterize DTD curves with a decreasing pattern (as the one reported in figure 2c) whereas smaller (or larger) PC2 scores indicate increasing patterns. This component was named SLOPE. The interpretation of LEVEL and SLOPE meaning may be better inferred from the average DTD curves for different PC1 and PC2 classes (Figures 2). Only MY data were reported for brevity, but the other traits showed the same pattern.

To simplify the comparison between SLOPE for different traits, scores of SLOPE_PP were multiplied by -1. Pearson correlations among PC scores (Table 2) confirm the expected orthogonality between LEVEL and SLOPE within each trait. Sign and magnitude of correlations between PC scores of different traits confirm the meaning of the new variables extracted: examples are the negative correlations between LEVEL for milk yield and for fat and protein percentage and the positive correlation between the last two traits.

Genetic parameter estimation

All new variables exhibited genetic variability (Table 3). In particular, pseudo-heritability was moderate to high for LEVEL across all traits (values ranged between 0.32 for SCS and 0.82 for FP, respectively) and low to moderate for SLOPE (range from 0.16 for SCS to 0.28 for PP, respectively). SLOPE_MY has pseudo h^2 similar to heat tolerance measures estimated from milk yield data using different approaches (Nguyen et al., 2016; Sanchez et al.,2009). Moreover, values for SLOPE_FP and SLOPE_PP are consistent with the proportion of variability explained by one of the canonical variables obtained for the eigendecomposition of the additive (co)variance matrix of random regression coefficients for these traits (Carabano et al., 2014). Pseudo genetic correlations confirm also at genetic level the substantial orthogonality between the LEVEL and SLOPE components within each trait. Values of pseudo r_g between LEVEL values across the different traits were large (absolute values >0.65) with the exception of comparisons involving SCS. Large pseudo genetic correlations were also observed between SLOPE values for all traits, with the exception of the correlation between FP and SCS (Table 3).

Genome wide association study

Eight SNP were significantly associated to the considered traits (FDR<0.10; P_Bonf<0.08) (Table 4). Two of them were associated to SLOPE_MY. Four, including the three top significant SNP, were associated to LEVEL for FP and two for PP (one to LEVEL and one to SLOPE). No significant association was found for principal components extracted from DTD for SCS.

Milk yield. The two markers significantly associated to DTD_MY were both related to SLOPE (Table 4 and Figure 3a), the principal component that expresses the shape of individual curves for increasing levels of THI. The first SNP was located on BTA26 at approximately 22.3 Mb. A possible candidate gene located within the interval defined by this marker is the beta-transducin repeat containing E3 ubiquitin protein ligase (***BTRC***) (Table 5), reported to be associated with milk production (Raven et al., 2016) and leg morphology (van den Berg et al., 2014) in cattle, and with growth rate in chicken (Zhang et al., 2015). The 0.5 Mb windows includes also genes involved in folliculogenesis in cattle, as fibroblast growth factor 8 (***FGF8***), found in a selection sweep study in dairy cattle (Kemper et al., 2014), meningioma expressed antigen 5 (hyaluronidase) (***MGEA5***), and taurusKv channel interacting protein 2 (***KCNIP2***) (Hatzirodos et al., 2014). Of interest is also Hermansky-Pudlak syndrome 6 (***HPS6***), a gene related to pigmentation in humans (Sturm and Luy, 2012). A marker located on BTA26 and significantly associated to sweating rate was reported by Dikmen et al. (2013) for Holstein, although the map position is about 2.0 Mb far from the SNP identified in our study.

The second marker associated to SLOPE for milk yield was located on BTA6 at approximately 35.5 Mb (Table 4). In this region maps the coiled-coil serine rich protein 1 (***CCSER1***) (Table 5), a gene involved in the mechanism of cell division found to be associated to birth weight in Brangus cattle (Saatchi et al., 2014) and with Na concentration in muscle of Nelore cattle (Tizioto et al., 2015). Dikmen et al. (2013; 2015) found markers significantly associated to rectal temperature and

respiration rate on BTA6, but at a position about 10Mb far from the marker flagged in the present study.

Fat percentage. The four significant SNP detected for LEVEL of FP were located on BTA14 in an interval of approximately 0.6 Mb (Table 4). The top significant SNP (ARS-BFGL-NGS-4939) was reported to be significantly associated to milk fat also in Italian, US and German Holsteins (Cole et al., 2011; Wang et al., 2014; Capomaccio et al., 2015), in Italian Simmental cattle (Macciotta et al., 2015), and in a multibreed population (Raven et al., 2014). The 0.5 Mb window surrounding this top marker flagged a zone characterized by a high density of annotated genes. In particular in that region maps one of the most important genes affecting milk fat content and yield in cattle (Table 5), the DiacylAcyilGlicerolTransferase 1 (***DGAT1***) (Grisart et al., 2002). However, on BTA14 at approximately 1.81-1.83 Mb maps the heat shock transcription factor 1 (HSF1), a protein that is involved in the mechanism of response to heat stress (Guettouche et al., 2005). Among the genes that map in the interval defined by the second marker (Table 4), of interest there is the Rho GTPase activating protein 39 (***ARHGAP39***), involved in the development of the central nervous system (Ma and Novak, 2011) (Table 5). This gene has been reported to be associated to milk fat composition in Danish (Buitenhuis et al., 2014) and North American (Nayeri et al., 2016) Holstein, and to somatic cell score (Wang et al., 2015) in Chinese Holstein. Another interesting gene located in this region is the ribosomal protein L8 (***RPL8***), involved in the cellular mechanisms of homeostasis (Katz et al., 2016), associated to response to acute heat stress in the fish *Latescalcarifer* (Newton et al., 2012).
The third marker located on BTA14 pointed out the mitogen-activated protein kinase 15 (***MAPK15***) (Table 5), suggested as a candidate gene for Somatic Cell Score in a study on Chinese Holstein (Wang et al., 2015). The last significant SNP for PC1 of DTD_FP was also located on BTA 14, very close to the second significant SNP. Close to this marker maps the zinc finger protein 34 (***ZNF34***), found to be associated to milk fat percentage in Chinese Holstein (Jiang et al., 2014) and

311 Italian Simmental (Macciotta et al., 2015). No significant SNP were detected for the SLOPE
312 component for fat percentage (Figure 3b)
313
314 ***Protein percentage.*** A SNP significantly associated with SLOPE_PP was found on BTA5, at
315 approximately 114.8 Mb (Table 4 and Figure 3c). In the 0.5 Mb interval flanking this marker is
316 located an interesting gene, the malonyl-CoA-acyl carrier protein transacylase (***MCAT***) (Table 5)
317 expressed in the mitochondrion and involved in fatty acid metabolism. This gene has been found to
318 be differentially expressed in chicken embryos exposed to heat challenge (Loyau et al., 2016).
319 Other genes of interest located in this interval are the sorting and assembly machinery component
320 (***SAMM50***), a mitochondrial protein found associated with serum triglyceride levels in humans
321 (Kitamoto et al., 2013); and the translocator protein (**TSPO**), a gene upregulated in atretic bovine
322 follicles (Hatzirodos et al., 2014).
323 The other significant marker associated with PP was related to LEVEL. It was located on BTA14,
324 in a region where no annotated genes have been retrieved in the UCSC genome database.
325
326 **DISCUSSION**
327
328 Traits able to describe efficiently the response of animals to heat stress are rather
329 problematic to be routinely measured in dairy cattle populations involved in breeding programs.
330 Dikmen et al. (2012) estimated that 13 to 17% of the variation in rectal temperature in cows during
331 heat stress is due to genetic differences. However, it will not be practical to select cows for heat
332 tolerance based on rectal temperature directly because this trait is not recorded on dairy farms. On
333 the contrary, patterns of milk production traits can be conveniently modeled to estimate tolerance to
334 heat stress (Carabano et al., 2014¸ Bernabucci et al., 2014; Hammami et al., 2015). In particular
335 DTD have been proposed as a proxy of individual response of bulls across increasing levels of heat
336 load (Hayes et al., 2009). The reliability of heat tolerance proxies based on production traits was

337 underlined in a recent study on dairy cattle subjected to heat challenge under controlled
338 environment, namely climate chambers (Garner et al., 2016). Heat tolerant cows, ranked according
339 to a genomic breeding value of heat tolerance based on milk yield, showed lower values of
340 physiological indicators (core body temperature, rectal and intra-vaginal temperature) than heat
341 sensitive cows.

342 In the present work DTD calculated for four different milk production traits were analyzed
343 with PCA. This approach was able to extract two new variables, which explained approximately
344 50% of the original variance. Interestingly, they were related to the level of the DTD curve and to
345 its slope, respectively. Our interpretation of PC meaning is in agreement with the outcome of eigen
346 decomposition of coefficient matrix of Random Regression models used to estimate heat tolerance
347 (Carabano et al., 2014) and with previous reports on PCA carried out on milk production traits in
348 cattle (Macciotta et al., 2006; 2015).

349 Most of genetic models used to study the heat stress effect on dairy traits use the reaction
350 norm model approach (Kormodin and Bijma, 2004; Shariati et al., 2007). In particular, for each
351 animal the THI effect is fitted as a general intercept plus a slope (Carabano et al., 2014; Hayes et
352 al., 2009; Sanchez et al., 2009). The results of the model free PCA approach used in the present
353 study basically confirm theoretical assumptions of the reaction norm model. Also, the
354 predominance of LEVEL over SLOPE in terms of variance explained is in agreement with previous
355 reports. In particular, SLOPE eigenvalue (about 10%) is not far from values reported for the second
356 eigenfunction for fat and protein yield obtained from the decomposition of (co)variance matrix of
357 random regression models by Carabano et al. (2014). These evidences suggest that the behavior of
358 milk production traits across increasing THI levels can be partitioned into two main components,
359 one basically related to the overall production genetic potential of the animal and the other to the
360 individual specific response.

361 The SLOPE component could be proposed as a measure of individual tolerance to heat stress.
362 However, the interpretation of its values deserves further discussion. From figures 3b it can be

363 noticed that animals having negative SLOPE scores exhibited, on average, lower DTD in the first
364 part of the THI scale (i,e. in the comfort zone) and higher in the second part, respectively. Thus a
365 selection in favor of this kind of response pattern would result in less productive animals under
366 comfort conditions, i.e. the most frequent at least in temperate areas. However, it should be
367 remembered that graphs depicted in figures 2 represents deviations from the average response
368 pattern (Carabano et al., 2016). Thus animals exhibiting a curve as the one reported in figure 2b are
369 not expected to respond with a decrease of production to lower THI levels but to have a quite
370 constant average level of production in this point of the curve. Moreover, provided the
371 orthogonality between the two traits, the crossed distribution of individual patterns among the
372 classes of the two PCs (Table 6) shows values in all the cells. Although intermediate classes (i.e.
373 those having values between -1 and 1 for both the components) are the most abundant there are 10
374 bulls that belong to the class having the most positive and negative values for LEVEL and SLOPE
375 respectively. The DTD_MY patterns of some of these bulls are reported in figure 4. They exhibit
376 the typical ascending pattern of this SLOPE class, with the points in most part having values higher
377 than zero or showing great variability.

378 Another point of agreement between the results of the present study and those of other studies
379 on heat tolerance based on the broken line model, can be found in the structure of PC eigenvectors.
380 It is worth noticing that the inversion of eigenvector coefficient sign of SLOPE (Table 1) occurs in
381 the THI class 68-70 for DTD of MY, PP and SCS, and in the class 65-67 for FP respectively. These
382 values basically agreed with estimates of THI_0 threshold reported by some authors. Sanchez et al.
383 (2009) estimated a THI threshold of 72 for milk yield in US Holsteins using hierarchical models.
384 Carabano et al. (2016) estimated THI_0 thresholds of 72-73 for milk yield in Holstein populations of
385 different European countries. Also the lower threshold of fat percentage compared with the other
386 traits reported in the present study agrees with previous reports. In our previous study a THI
387 thresholds of 76, 73 and 74 were estimated in first, second, and third parity Holstein cows for MY
388 (Bernabucci et al., 2014).

In view of a possible implementation of PCA-derived measures of tolerance to heat stress in breeding programs, the two stage approach suggested in this study has the limitation that only bulls having complete records (i.e. the 11 points of the DTD curve) could be considered. Such a requirement strongly reduced the number of animals for which PC could be computed (for example from 4,184 to 1,540 for milk yield in the present study). To overcome this problem, the correlation matrix between different points of the DTD curve could be reconstructed by using a random regression model (**RRM)** in which records at different THI classes are treated as repeated measures for each sire. Thus DTD_MY were fitted with a mixed model having the same structure of [1] but with the sire effect fitted as random. The covariance within animal was accounted for by an 11x11 unstructured matrix of between sire effects for each THI level. To avoid convergence problems, bulls having 7 or more records were considered. A total of 35,992 records belonging to 3,697 bulls were used. From table 7, it can be seen that observed Pearson correlations among DTD_MY and the correlation matrix estimated by the RRM are very similar. Moreover, also the plot of observed and estimated correlations averaged for the lag in THI class (Figure 5) underscores the substantial equivalence between the two approaches Finally, PCA carried out on the RRM estimated correlation matrix yielded basically the same results of the two-step approach, both in terms of variance explained by the first two PC and of their eigenvector structure (i.e., the first can be considered a LEVEL PC, the second a SLOPE, respectively). Such a concordance of results opens interesting perspectives on the possible use of this heat tolerance indices on large scale.

Previous studies on the genetic basis of tolerance to heat stress indicated that this trait has a moderate genetic variability and that it could be split into an intercept and a slope component that are genetically related. The PC based measure of heat tolerance proposed in the present work does show genetic variability. In particular, the pseudo heritability of the SLOPE component is similar to h^2 values reported in other studies for the slope parameter of the reaction norm model. Thus the two PC extracted from DTD of four milk traits could be considered as possible breeding criteria when selecting for improving heat tolerance in cattle. However, compared to previous measurements of

tolerance to heat stress based on milk production data, a distinguishing feature of the PC is their phenotypic and genetic orthogonality. The absence of any genetic relationship between LEVEL and SLOPE (Table 2) suggests that an independent selection of the two main aspects of DTD patterns, i.e. level of production and heat tolerance, should be feasible. A simultaneous section for improving both heat tolerance and dairy traits could be achieved also by implementing a selection index in which suitable economic weight have to be determined (Nguyen et al., 2016). However, provided the unfavorable genetic correlation between heat tolerance and production, a smaller selection response is expected for each single trait in comparison with the use of the two uncorrelated LEVEL and SLOPE component.

The amount of variance accounted for by the slope parameter in reaction norm models is used as an indicator of the genotype x environment interaction (Kolmodin e Bijma; 2004; Shariati et al., 2007). In the present study, SLOPE /LEVEL additive variance ratios were 0.10 for MY, 0.05 for FP, 0.06 for PP, and 0.13 for SCS, respectively. These values confirm results of Santana et al. (2016) that concluded that G X E due to heat stress is more relevant for milk yield and somatic cell score than for fat and protein percentage.

Association analysis highlighted a limited number of significant markers. This is not an unexpected outcome, due to the severe correction needed to account for multiple testing when high throughput platforms are used, to the complex biology underlying the physiological response to heat stress, and to the limited size of the sample of animals genotyped.

Half of significant SNPs were associated to the LEVEL variable for FP and located on BTA14. This result is quite common in genomic studies on Holstein cattle, mainly due to the genetic architecture of the trait, which is largely influenced by a single segregating gene, the *DGAT1* (Grisart et al., 2002). The relevant influence of this gene on FP was previously observed in the Italian Holstein population (Fontanesi at al., 2014).

However, it is worth mentioning that a gene encoding for a heat shock transcription factor (***HSF1***) maps on BTA14 very close to the top significant marker detected in the present study. In

Humans *HSF1* mediates the expression of the heat shock and stress proteins in response to physical and chemical stresses (Guettoche et al., 2005). This gene has been recognized to have a central role in coordinating thermal tolerance in cattle (Collier et al., 2008). Differences in the expression of *HSF1* in the liver were found between cows calving in spring and summer, respectively (Shahzad et al., 2015). Moreover, associations between this gene and thermo-tolerance have been detected in Chinese Holstein (Li et al., 2011). An overexpression of *HSF1* has been found in Buffalo during summer under tropical environment (Kumara et al., 2015) and reported to be associated to genetic susceptibility to *Mycobacterium bovis* infection in dairy cattle (Richardson et al., 2016). Differential expression of heat shock proteins (***HSP***) has been related to *in vitro* fertilization rate and blastocyst rate of bovine embryos (Zhang et al., 2011). *HSP70.1* polymorphism has been associated to cellular thermo-tolerance in Holstein lactating cows (Basiricò et al., 2011)

The relationships between tolerance to heat stress and fertility has been confirmed by associations found in the present study Candidate genes found in the 0.5 Mb intervals defined by the significant markers for the SLOPE component are involved in cellular regulation mechanism, fertility, and weight at birth. Nguyen et al. (2016), using the DTD deviation as indicators of heat stress in Australian Holsteins, found favorable correlation between DTD and fertility.

Among putative candidate genes that have been detected in the present study for the SLOPE trait, one (***MCAT***) is involved in the fatty acid (**FA**) metabolism. This result is in agreement with findings by Hammami et al. (2015) that highlighted a relationship between milk FA content and tolerance to heat stress in cattle. It should pointed out that milk FA profile is an index of animal energy balance (Bastin et al., 2011) and it is strongly related to the diet, which may be affected by climatic conditions. Nardone et al. (1997) found greater proportions of long-chain FA in colostrum produced by heifers under heat stress conditions. Those authors demonstrated that the higher proportion of long-chain FA was due to the reduced synthesis of short- and medium-chain FA in the mammary gland cells. On the basis of all these evidences a role of milk FA as potent biomarkers for evaluating individual thermo tolerance could be hypothesized.

Previous GWAS studies carried out on DTD and rectal temperature have highlighted genomic regions associated with heat stress tolerance. These results were not confirmed in the present study, even though significant markers were evidenced for SLOPE for milk yield on BTAs 5, 6, and 26, i.e. on the same chromosomes where Dikmen et al. (2013) found significant associations for rectal temperature. The detection of a limited number of significant markers and a poor repeatability of results across studies and populations is a major issue in GWAS studies carried out on livestock species. Sample size, genetic differences among populations (i.e., level of linkage disequilibrium, allelic SNP frequencies) are mentioned as main reasons for such a lack of concordance between experiments. Moreover, the severe corrections of significance levels due to huge number of repeated tests further reduce the number of detected markers. Finally, a more general issue that has been raised in GWAS carried out in humans is that not all the genetic variation of a trait is captured by available markers, i.e. the so called problem of "missing heritability" (Gusev et al., 2013). On the other hand, in the case of GWAS for heat tolerance traits the role of the phenotypes should be carefully considered. Dikmen et al (2015) found that 1 out of 4 significant SNP previously detected using a phenotype derived from milk yield was also associated to a physiological indicator of heat stress (/i.e., sweating rate). However, SNP validation is problematic also within the same trait. Hayes et al. (2009), on a second independent data set, confirmed only 2 out of 42 SNP significantly associated to the slope component of a reaction norm model fitted to milk yield in a Holstein population. A further issue in genetic studies of tolerance to heat stress is represented by the methodologies used to obtain the environmental variable. THI calculations, for example, could differ in the kind of variable used (i.e. daily maximum or average temperature, minimum or average relative humidity) and in the time lag with the day of the test. Such an heterogeneity of measures could be one of the reasons for the differences between studies in the estimates of the THI upper threshold for the comfort zone. The existence of all these sources of variation that may possibly affect results of studies on the genetic dissection of heat tolerance traits should lead scientists to make efforts in increasing the power of their experiments, validating results in independent

populations, and in harmonizing methodologies for calculating environmental variables, Furthermore, alternatives to traditional methods for measuring physiological indicators of heat tolerance should be found. Recent achievements in precision farming as the use of micro-equipment directly installed on the animal or the use of indirect predictors of physiological traits (as for example MIR milk spectra) could provide interesting tools.

CONCLUSIONS

The PCA was able to derive two new variables able to describe the overall level and the slope of response of milk production traits across increasing levels of THI index. These two new phenotypes are uncorrelated and therefore may provide an option for overcoming the problem of the negative correlation between heat tolerance and production level that has been found within the context of milk and weather record analysis. The genetic background of the two PC was investigated and some putative candidate genes have been proposed. A useful numerical property of the two extracted variables is their orthogonality. This feature makes the use of the second PC as a measure of thermo-tolerance for breeding purpose particularly appealing because many authors have stressed the need for using measures of tolerance to heat stress uncorrelated from the production level. Moreover, this variable could be derived from data that are routinely recorded in breeding programs and therefore its use on large scale could be proposed.

ACKNOWLEDGEMENTS

The research was funded by the Italian Ministry of Agriculture (SELMOL grant). Authors thank the Italian Holstein Association (ANAFI) for providing milk production and pedigree data.

REFERENCES

Aguilar, I., I. Misztal, S. Tsuruta. 2009. Genetic components of heat stress for dairy cattle with multiple lactations. J.Dairy Sci. 92:5702-5711

519 Aguilar, I., I. Misztal, S. Tsuruta. 2010. Short communication: Genetic trends of milk yield under
520 heat stress for US Holsteins. J. Dairy Sci. 93:1754-1758.
521 Ali, A.K.A., and G.E. Shook. 1980. An optimum transformation for somatic cell concentration in
522 milk. J. Dairy Sci. 63: 487-490.
523 Amin, N., C.M. van Duijn, and Y.S.Aulchenko. 2007. A genomic background based method for
524 association analysis in related individuals. PLosONE 2:e1274.
525 Basiricò, L., P. Morera, V. Primi, N. Lacetera, A. Nardone, and U. Bernabucci. 2011. Cellular
526 thermotolerance is associated with heat shock protein 70.1 genetic polymorphisms in Holstein
527 lactating cows. Cell Stress Chaperones 16:441–448.
528 Bastin , C., N. Gengler , and H. Soyeurt. 2011. Phenotypic and genetic variability of production
529 traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. J.
530 Dairy Sci. 94 :4152–4163
531 Bernabucci, U., N. Lacetera, L. H. Baumgard, R. P. Rhoads, B. Ronchi B., Nardone A. 2010.
532 Metabolic and hormonal acclimation to heat stress in domesticated ruminants. Animal 4:1167-
533 1183.
534 Bernabucci, U., S. Biffani, L. Buggiotti, A. Vitali, N. Lacetera, and A. Nardone. 2014. The effect of
535 heat stress in Italian Holstein dairy cattle. J. Dairy Sci. 97: 471-486.
536 Biffani, S., U. Bernabucci, A. Vitali, N. Lacetera, A. Nardone.2016. Short communication: Effect
537 of heat stress on nonreturn rate of Italian Holstein cows. J. Dairy Sci. 99: 5837-5843.
538 Bohmanova, J., I. Misztal, S. Tsuruta, H.D. Norman, T.J. Lawlor. 2008. Short Communication:
539 Genotype by Environment Interaction Due to Heat Stress. J. Dairy Sci. 91:840-846.
540 Brügemann , K., E. Gernand , U. U. von Borstel , and S. König. 2012. Genetic analyses of protein
541 yield in dairy cows applying random regression models with time-dependent and temperature x
542 humidity-dependent covariates. J. Dairy Sci. 94 :4129–4139.

543 Buitenhuis, B., L.L.G. Janss, , N. A. Poulsen, L. B. Larsen, M. K. Larsen, and P. Sørensen. 2014.
544 Genome-wide association and biological pathway analysis for milk-fat composition in
545 DanishHolstein and Danish Jersey cattle. BMC Genomics 15:1112.
546 Carabano, M.J., K. Bachaga, M. Ramon, and C. Diaz. 2014. Modeling heat stress effect on Holstein
547 cows under hot and dry conditions: selection tools. J. Dairy Sci. 97:7889-7904.
548 Carabano, M.J., B. Logar, J. Bormann, J. Minet, M.L. Vanrobays, C. Diaz, B. Tychon, N. Gengler,
549 and H. Hammami. 2016. Modeling heat stress under different environmental conditions. J.
550 Dairy Sci. 99:3798–3814.
551 Capomaccio, S., M. Milanesi, L. Bomba, K. Cappelli, E. L. Nicolazzi, J. L. Williams, P. Ajmone-
552 Marsan, and B. Stefanon. 2015. Searching new signals for production traits through gene-based
553 association analysis in three Italian cattle breeds. Animal Genetics 46:361-370.
554 Cole, J.B., G.R. Wiggans, L. Ma, T.S. Sonstegard, T.J. Lawlor Jr., B.A. Crooker, C.P. Van Tassell,
555 J. Yang, S. Wang, L.K. Matukumalli, and Y. Da. 2011.Genome-wide association analysis of
556 thirty one production, health, reproduction and body conformation traits in contemporary
557 U.S.Holstein cows. BMC Genom. 12:408.
558 Collier, R. J. ,J. L. Collier, R. P. Rhoads, and L. H. BaumgardJ. 2008. Invited Review: Genes
559 Involved in the Bovine Heat Stress Response Dairy Sci. 91:445–454
560 Fontanesi, L., D. G. Calò, G. Galimberti, R. Negrini, R. Marino, A. Nardone, P. Ajmone-Marsan,
561 and V. Russo. 2014. A candidate gene association study for nine production traits in Italian
562 Holstein sires. Animal Genetics 45:576-580.
563 Dikmen, S., J. B. Cole , D. J. Null , and P. J. Hansen. 2012. Heritability of rectal temperature and
564 genetic correlations with production and reproduction traits in dairy cattle. J. Dairy Sci. 95
565 :3401–3405.
566 Dikmen, S., J. B. Cole , D. J. Null , and P. J. Hansen. 2013. Genome-wide association mapping for
567 identification of quantitative trait loci for rectal temperature during heat stress in Holstein
568 cattle. PLosONE 8:e69202.

569 Dikmen, S., X.Z. Wang, M.S. Ortega, J.B. Cole, D.J. Null, and P.J. Hansen. 2015. Single nucleotide
570 polymorphisms associated with thermoregulation in lactating dairy cows exposed to heat stress.
571 J. Anim. Breed. Genet. 132: 409–419
572 Garner, J.B., M.L. Douglas, S.R.O Williams, W.J. Wales, L.C. Marett1, T.T.T. Nguyen,C.M.
573 Reich, and B. J. Hayes. 2016. Genomic Selection Improves Heat Tolerance in Dairy Cattle.
574 Scientific Reports DOI: 10.1038/srep34114.
575 Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid,
576 P. Simon, R. Spelman, M. Georges, and R. Snell. (2002) Positional candidate cloning of a QTL
577 in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major
578 effect on milk yield and composition. Genome Res. 12: 222-231.
579 Guettouche, T., Boellmann, W.S. Lane.And R. Voellmy. 2005. Analysis of phosphorylation of
580 human heat shock factor 1 in cells experiencing a stress. BMC Biochemistry 2005, 6:4
581 doi:10.1186/1471-2091-6-4.
582 Gusev, A., G. Bhatia, N. Zaitien, B. J. Vilhjalmsson, D. Diogo, E. A. Stahl, P. K. Gregersen, J.
583 Worthington, L. Klareskog, S. Rayachaudhri, R.M. Plenge, B. Pasanius, A.L. Price. 2013.
584 Quantifying missing heritability at known GWAS loci. PLoSGenetics e1003993.
585 Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bastin, N. Gengler. 2015.Genetic
586 analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein
587 cows. J. Dairy Sci. 98:4956-4968
588 Hansen, P. J. 2007. Exploitation of genetic and physiological determinants of embryonic resistance
589 to elevated temperature to improve embryonic survival in dairy cattle during heat stress.
590 Theriogenology 68:S242–S249.
591 Hatzirodos, N., H. F. Irving-Rodgers, K. Hummitzsch, M. L. Harland, S. E. Morris, and R. J.
592 Rodgers. 2014. Transcriptome profiling of granulosa cells of bovine ovarian follicles during
593 growth from small to large antral sizes. BMC Genomics 15:24.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain, K. Savin, C.P. van Tassell, T.S. Sostengard, and M.E. Goddard. 2009. A validated genome wide association study to breed cattle adapted to an environment altered by climate changes. PLosONE. 4:e6676.

Jiang, L., X. Liu, J. Yang, H. Wang, J. Jiang, L. Liu, S. He, X. Ding, J. Liu, and Q. Zhang. 2014. Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC Genomics 15:1105.

Kadzere, C.T , M.R. Murphy, N. Silanikove, and E. Maltz. 2002. Heat stress in lactating dairy cows: a review. Livest. Prod. Sci. 77: 59–91.

Katz, M. J., L. Gandara, A. L. De Lella Ezcurra, and P. Wappner. 2016. Hydroxylation and translational adaptation to stress: some answers lie beyond the STOP codon. Cell. Mol. Life Sci. 73:1881–1893

Kelly, C. F., and T. E. Bond. 1971. Bioclimatic factors and their measurements. Page 7 in A guide to environmental research in animals. Natl. Acad. Sci., Washington, DC.

Kemper, K.E., S.J. Saxton, S. Bolormaa, B.J. Hayes, , M.E. Goddard. 2014. Document Selection for complex traits leaves little or no classic signatures of selection. BMC Genom. 15 (1), 246

Kitamoto, T., A. Kitamoto, M. Yoneda, H. Hyogo, H. Ochi, T. Nakamura, H. Teranishi, S. Mizusawa, T. Ueno, K. Chayama, A. Nakajima, K. Nakao, A. Sekine, and K. Hotta. 2013. Genome-wide scan revealed that polymorphisms in the PNPLA3, SAMM50, and PARVB genes are associated with development and progression of nonalcoholic fatty liver disease in Japan. Hum Genet. 132:783–92.

Kolmodin, R., and P. Bijma. 2044. Response to mass selection when the genotype by environment interaction is modeled as a linear reaction norm. Genet. Sel. Evol. 36:435-454.

Kumara, A., S. Ashrafa, T. Sridhar Gouda, A. Grewalc, S.V. Singha, B.R. Yadavb, R.C. Upadhyaya. 2015. Expression profiling of major heat shock protein genes during different seasons in cattle (Bosindicus) and buffalo (Bubalusbubalis) under tropical climatic condition. J. Thermal Biol.51: 55–64.

620 Li, Q.L., Z.H. Ju, J.M. Huang, J.B. Li, R.L. Li, M.H Hou, C. F. Wang, and J.F.Zhong. 2011. Two
621 Novel SNPs in HSF1 Gene Are Associated with Thermal Tolerance Traits in Chinese Holstein
622 Cattle. DNA Cell Biol. 30:247-254.
623 Loyau, T., C. Hennequet-Antier, V. Coustham, C. Berri, M. Leduc, S. Crochet, M. Sannier, M. J.
624 Duclos, S. Mignon-Grasteau, S. Tesseraud, A. Brionne, S. Métayer-Coustard, M. Moroldo, J.
625 Lecardonnel, P. Martin, S. Lagarrigue, S. Yahav, and A. Collin. 2016 Thermal manipulation of the
626 chicken embryo triggers differential gene expression in response to a later heat challenge. BMC
627 Genomics17:329.
628 Ma, S., and F. V. Nowak. 2011. The RhoGAP domain-containing protein, Porf-2, inhibits
629 proliferation and enhances apoptosis in neural stem cells. Mol. and Cell. Neuroscience 46:573–
630 582].
631 Macciotta, N.P.P., D. Vicario, and A. Cappio-Borlino A. 2006. Use of multivariate analysis to
632 extract latent variables related to level of production and lactation persistency in dairy cattle. J.
633 Dairy Sci. 89: 3188-3194.
634 Macciotta, N. P. P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-
635 Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve
636 traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175–8185
637 Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related
638 programs (BGF90). Pages 743–744 in Proceedings of the 7th World Congress on Genetics
639 Applied to Livestock Production, Montpellier, France
640 Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F.Schenkel, S. S. Moore, and
641 P.Stothard. 2016. Genome-wide association for milk production and female fertility traits in
642 Canadian dairy Holstein cattle. BMC Genet. 17:75.
643 Nardone, A., N. Lacetera, U. Bernabucci, B. Ronchi. 1997. Composition of Colostrum from Dairy
644 Heifers Exposed to High Air Temperatures During Late Pregnancy and the Early Postpartum
645 Period. J. Dairy Sci. 80:838–844

646 Nardone A., B. Ronchi, N. Lacetera, M. S. Ranieri, and U. Bernabucci. 2010. Effects of climate
647 changes on animal production and sustainability of livestock systems. Liv. Sci. 130:57-69.
648 Newton, J. R., C. De Santis, and D. R. Jerry. 2012. The gene expression response of the
649 catadromous perciform barramundi Latescalcarifer to an acute heat stress. J.Fish Biol. 81: 81–
650 93.
651 Nguyen, T.T.T., P.J. Bowman, M. Haile-Mariam, J.E. Pryce, and B.J. Hayes. 2016. Genomic
652 selection for tolerance to heat stress in Australian dairy cattle. J. Dairy Sci. 99: 2849-2862
653 Perano, K. M., J.G. Usack, L.T. Angenent, K.G. Gebremedhin. 2015. Production and physiological
654 responses of heat-stressed lactating dairy cattle to conductive cooling. J. Dairy Sci. 98:5252–
655 5261.
656 Ravagnolo, O., and I. Misztal. 2000. Genetic Component of Heat Stress in Dairy Cattle, Parameter
657 Estimation. J. Dairy Sci. 83: 2126-2130Raven, L.A., B.G. Cocks, K.E. Kemper,
658 A.J.Chamberlain, M.E. Goddard, B.J. Hayes. 2016. Targeted imputation of sequence variants
659 and gene expression profiling identifies twelve candidate genes associated with lactation
660 volume, composition and calving interval in dairy cattle. Mammalian Genome 27: 81-97.
661 Richardson, I.W., D. P. Berry, H. L. Wiencko, I. M. Higgins, S. J. More, J. McClure, D. J. Lynn,
662 and D. G. Bradley. 2016. A genome-wide association study for genetic susceptibility to
663 *Mycobacterium bovis* infection in dairy cattle identifies a susceptibility QTL on chromosome
664 23. Genet. Sel. Evol. 48:19
665 Saatchi, M., R. D. Schnabel, J. F. Taylor, and D. J Garrick. 2014. Large-effect pleiotropic or closely
666 linked QTL segregate within and across ten US cattle breeds. BMC Genom., 15,442–458.
667 SAS Institute. 2008. SAS/STAT version 9.2. SAS Inst. Inc., Cary, NC.
668 Sánchez, J.P., I. Misztal, I. Aguilar, B. Zumbach, and R. Rekaya. 2009. Genetic determination of
669 the onset of heat stress on daily milk production in the US Holstein cattle. J. Dairy Sci. 92
670 :4035–4045.

671 Santana, M.L., A.B.Bignardi, R.J. Pereira, G. Stefani, and L. El Faro.2016. Genetics of heat
672 tolerance for milk yield and quality in Holsteins. Animal doi:10.1017/S1751731116001725.
673 Segnalini, M., A. Nardone, U. Bernabucci, A.Vitali, B. Ronchi, N. Lacetera, 2011. Dynamics of
674 the temperature-humidity index in the Mediterranean basin. Int. J. Biomet. 55: 253-263
675 Shariati, M.M., G. Su, P. Madsen, and D. Sorensen. 2007. Analysis of milk production traits in
676 early lactation using a reaction norm model with unknown covariates. J. Dairy Sci. 90: 5759-
677 5766.
678 Sturm, R.A., and D. L. Duy. 2014. Human pigmentation genes under environmental selection.
679 Genom. Biol. 2012, 13:248.
680 Tizioto, P.C., J.F. Taylor, J.E. Decker, C.F. Gromboni, M.A.Mudadu, R.D. Schnabel, L.L.Coutinho,
681 G.B. Mourão, P.S.N. Oliveira, M.M. Souza, J.M. Reecy, R.T. Nassu, F.A. Bressani, P. Tholon,
682 T.S.Sonstegard, M. M. Alencar, R.R.Tullio, A. R. A. Nogueira, and L.C.A.Regitano. 2015.
683 Detection of quantitative trait loci for mineral content of Nelore longissimus dorsi muscle.
684 Genet. Sel. Evol. 47:15.
685 Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea,
686 C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and
687 S. L.Salzberg. 2009. A whole-genome assembly of the domestic cow, Bostaurus. Genom. Biol.
688 10:R42.
689 van den Berg, I., S. Fritz, S. Rodriguez, D. Rocha, M. Boussaha, M. S. Lund and D. Boichard.
690 2014. Concordance analysis for QTL detection in dairy cattle: a case study of leg morphology.
691 Genetics Selection Evolution 46:31.
692 Shahzad,K., H. Akbar, M. Vailati-Riboni, L. Basiricò, P. Morera, S. L. Rodriguez-Zas, A.Nardone,
693 U. Bernabucci,and J. J. Loor. 2015. The effect of calving in the summer on the hepatic
694 transcriptome of Holstein cows during the peripartal period.J. Dairy Sci. 98:5401–5413

695 Zhang, B. F.,Peñagaricano, A. Driver , H. Chen , and H. Khatib. 2011.Differential expression of
696 heat shock protein genes and their splice variants in bovine preimplantation embryos. J. Dairy
697 Sci. 94 :4174–4182.
698 Zhang, G.X., Q.C. Fan, T. Zhang, J.Y. Wang, W.H. Wang, Q. Xue, and Y.J. Wang. 2015. Genome-
699 wide association study of growth traits in the Jinghai Yellow chicken. Genet. Mol. Res. 14:
700 15331-15338.
701 Wang, X., C. Wurmser, H. Pausch, S. Jung, F. Reinhardt, J. Tetens, G. Thaller, and R. Fries. 2014.
702 Identification and dissection of four major QTL affecting milk fat content in the German
703 Holstein-Friesian population. PLoS ONE 7:e40711.
704 Wang, X., P. Ma, J. Liu, Q. Zhang, Y. Zhang, X. Ding, L. Jiang, Y. Wang, Y. Zhang, D. Sun, S.
705 Zhang, G. Su and Y. Yu. 2015. Genome-wide association study in Chinese Holstein cows
706 reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility.
707 BMC Genetics 16:111.
708
709

Table 1. Eigenvectors and eigenvalues of the first two principal components (LEVEL and SLOPE) extracted from the correlation matrix of Daughter Traits Deviations for milk yield (DTD_MY), fat (DYD_FP), and protein (DTD_PP) percentages, and Somatic cell score (DTD_SCS)

THI interval	DTD_MY		DTD_FP		DTD_PP		DTD_SCS	
	LEVEL	SLOPE	LEVEL	SLOPE	LEVEL	SLOPE	LEVEL	SLOPE
50-52	0.25	0.45	0.24	0.76	0.26	-0.48	0.22	0.49
53-55	0.28	0.41	0.29	0.37	0.29	-0.37	0.25	0.55
56-58	0.32	0.22	0.30	0.17	0.31	-0.20	0.33	0.07
59-61	0.31	0.14	0.31	-0.25	0.32	-0.07	0.30	0.12
62-64	0.32	0.17	0.31	0.03	0.32	-0.18	0.34	0.09
65-67	0.32	0.00	0.31	-0.02	0.33	-0.06	0.30	0.01
68-70	0.32	-0.04	0.32	-0.14	0.31	0.10	0.34	-0.11
71-73	0.32	-0.13	0.32	-0.28	0.31	0.22	0.31	-0.06
74-76	0.31	-0.31	0.31	-0.02	0.31	0.11	0.33	-0.26
77-79	0.30	-0.34	0.31	-0.30	0.29	0.36	0.32	-0.14
>79	0.25	-0.55	0.28	-0.11	0.25	0.60	0.27	-0.57
Eigenvalue	4.43	0.96	4.25	0.91	4.79	0.91	3.60	0.98
Eigenvalue %	40	9	39	8	43	8	33	9

Table 2. Pearson correlations between the scores of the first principal components extracted from the correlation matrix of Daughter Traits Deviations for milk yield (MY), fat (FP) and protein (PP) percentage, and somatic cell score (SCS)

	LEVEL MY	SLOPE MY	LEVEL FP	SLOPE FP	LEVEL PP	SLOPE PP	LEVEL SCS	SLOPE SCS
LEVEL MY	1.00	0.00	-0.37	-0.03	-0.34	0.00	-0.07	-0.07
SLOPE MY		1.00	0.01	-0.12	-0.01	-0.20	-0.04	-0.16
LEVEL FP			1.00	0.00	0.52	0.02	-0.05	0.04
SLOPE FP				1.00	0.03	0.23	-0.02	0.05
LEVEL PP					1.00	0.00	0.05	0.02
SLOPE PP						1.00	0.02	-0.15
LEVEL SCS							1.00	0.00
SLOPE SCS								1.00

Table 3. Pseudo-heritability (on the diagonal) and pseudo-genetic correlations (out of the diagonal) between the scores of the first principal components extracted from the correlation matrix of Daughter Traits Deviations for milk yield (MY), fat (FP) and protein (PP) percentage, and somatic cell score (SCS). Standard errors of pseudo-heritabilities are reported within brackets.

	LEVEL MY	SLOPE MY	LEVEL FP	SLOPE FP	LEVEL PP	SLOPE PP	LEVEL SCS	SLOPE SCS
LEVEL MY	0.52 (.07)	0.02	-0.67	-0.05	-0.67	0.01	-0.05	-0.19
SLOPE MY		0.24 (.04)	0.04	-0.56	-0.05	-0.83	-0.17	-0.85
LEVEL FP			0.82 (.11)	0.01	0.75	0.04	-0.10	0.14
SLOPE FP				0.23 (.01)	0.06	0.88	-0.10	0.36
LEVEL PP					0.74 (.11)	0.00	0.11	0.10
SLOPE PP						0.28 (.01)	-0.07	0.69
LEVEL SCS							0.32 (.07)	0.04
SLOPE SCS								0.16 (.05)

Table 4. Table 2. Markers significantly associated (FDR<0.10) with scores of principal components extracted from Daughter Trait Deviations for milk yield, fat and protein contents.

SNP	BTA	Position bp	P-GC_P Bonf[1]	FDR	Trait
ARS-BFGL-NGS-4939	14	1,801,116	0.000001	0.000001	LEVEL_FP
ARS-BFGL-NGS-57820	14	1,651,311	0.000021	0.000011	LEVEL_FP
ARS-BFGL-NGS-107379	14	2,054,457	0.000335	0.000112	LEVEL_FP
ARS-BFGL-NGS-29678	26	22,383,645	0.000973	0.000973	SLOPE_MY
Hapmap30383-BTC-005848	14	1,489,496	0.049244	0.012311	LEVEL_FP
ARS-BFGL-NGS-19275	5	114,818,206	0.053830	0.053830	SLOPE_PP
Hapmap32110-BTA-153952	6	35,555,247	0.077947	0.038974	SLOPE_MY
Hapmap32435-BTC-012188	14	56,075,435	0.081155	0.081155	LEVEL_PP

[1] = level of significance of the test adjusted with Bonferroni correction

Table 5. Putative candidate genes located in the 0.5 Mb interval surrounding the significant SNPs.

BTA	Position (Mb)	Gene Name and symbol	Trait	Function	Associations with phenotypic traits
26	22.09-22.17	beta-transducin repeat containing E3 ubiquitin protein ligase - *BTRC*	SLOPE_MY	Intracellular protein degradation	Milk production Leg morphology Growth rate
26	22.37-22.38	fibroblast growth factor 8 - *FGF8*	SLOPE_MY	Mitogenic and cell survival	Folliculogenesis
26	22.39-22.42	meningioma expressed antigen 5 (hyaluronidase) - *MGEA5*	SLOPE_MY	Protein glycosilation	Folliculogenesis
26	22.42-22.43	taurusKv channel interacting protein 2 - *KCNIP2*	SLOPE_MY	Ion transportation	Folliculogenesis
26	22.62-22.63	Hermansky-Pudlak syndrome 6 - *HPS6*	SLOPE_MY	Melanosome development	Pigmentation
6	35.10-35.95	coiled-coil serine rich protein 1 - *CCSER1*	SLOPE_MY	Cell division	Beef traits
14	1.79-1.81	DiacylAcyilGlicerolTransferase 1 - *DGAT1*	LEVEL_FP	Fat metabolism	Milk fat content
14	1.81-1.83	heat shock transcription factor 1 - *HSF1*	LEVEL_FP	Response to cell stress	Heat stress tolerance *Mycobacterium bovis* susceptibility
14	1.56-1.60	Rho GTPase activating protein 39 - *ARHGAP39*	LEVEL_FP	Nervous system development	Milk fat content
14	1.50-1.51	ribosomal protein L8 - RPL8	LEVEL_FP	Homeostasis	Heat stress (fish)
14	2.23-2.24	the mitogen-activated protein kinase 15 - *MAPK15*	LEVEL_FP	Cell proliferation	Somatic cell score
14	1.49-1.50	zinc finger protein 34 - *ZNF34*	LEVEL_FP	Transcription regulation	Milk fat
5	11.45-11.46	malonyl-CoA-acyl carrier protein transacylase - *MCAT*	SLOPE_PP	Fatty acid metabolism	Heat stress (chicken)
5	11.49-11.50	sorting and assembly machinery component - *SAMM50*	SLOPE_PP	Mitochondrial protein	Triglyceride levels (Humans)
5	11.45-11.46	Translocator protein - TSPO	SLOPE_PP	Cholesterol transportation	Folliculogenesis

Table 6. Absolute frequencies of individual patterns of Daughter Traits Deviations for milk yield across different classes of LEVEL and SLOPE PC scores.

		SLOPE					
		≤-2	-2 to -1	-1 to 0	0 to 1	1 to 2	≥2
LEVEL	≤-2	10	24	72	81	21	6
	-2 to -1[1]	4	15	82	80	25	4
	-1 to 0	5	32	138	137	31	4
	0 to 1	4	27	128	137	28	3
	1 to 2	7	21	90	84	16	8
	≥2	10	32	70	62	27	15

[1] = lower limit of the class is included;

Table 7. Pearson correlations among Daughter Trait Deviations for milk yield (above the diagonal) and estimated unistructured correlation matrix estimated with random regression model (under the diagonal)

	TH11	TH12	TH13	TH14	TH15	TH16	TH17	TH18	TH19	TH1110	TH111
TH11	*	0.41	0.37	0.35	0.36	0.37	0.33	0.29	0.31	0.30	0.23
TH12	0.43	*	0.42	0.40	0.40	0.40	0.37	0.36	0.35	0.33	0.25
TH13	0.38	0.44	*	0.43	0.46	0.44	0.43	0.41	0.43	0.36	0.26
TH14	0.35	0.41	0.43	*	0.41	0.44	0.42	0.42	0.40	0.34	0.33
TH15	0.35	0.40	0.44	0.41	*	0.44	0.44	0.41	0.43	0.37	0.30
TH16	0.34	0.39	0.42	0.42	0.42	*	0.44	0.44	0.42	0.43	0.35
TH17	0.32	0.37	0.41	0.41	0.43	0.40	*	0.43	0.45	0.43	0.34
TH18	0.28	0.35	0.38	0.40	0.37	0.41	0.39	*	0.42	0.41	0.43
TH19	0.27	0.30	0.35	0.36	0.39	0.39	0.44	0.42	*	0.45	0.37
TH110	0.23	0.26	0.33	0.27	0.31	0.37	0.43	0.42	0.47	*	0.41
TH111	0.18	0.22	0.22	0.26	0.26	0.34	0.34	0.41	0.41	0.43	*

Table 8. Eigenvectors and eigenvalues of the first two principal components extracted from the correlation matrix estimated by fitting a random regression model to Daughter Traits Deviations for milk yield (DTD_MY).

THI interval	PC1	PC2
50-52	0.26	-0.40
53-55	0.30	-0.36
56-58	0.31	-0.27
59-61	0.31	-0.21
62-64	0.31	-0.18
65-67	0.32	-0.03
68-70	0.32	0.07
71-73	0.31	0.19
74-76	0.31	0.30
77-79	0.29	0.42
>79	0.27	0.51
Eigenvalue	4.66	1.15
Eigenvalue %	42	10

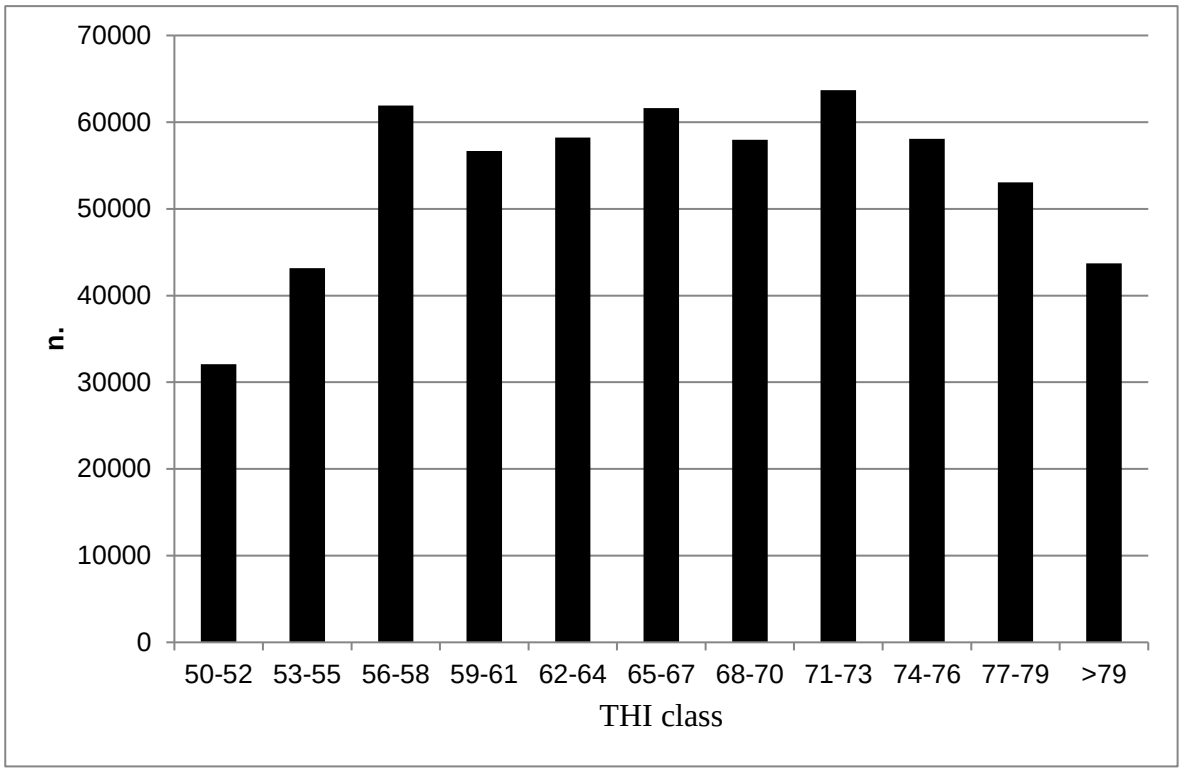

70000
60000
50000
40000
30000
20000
10000
0
n.
50-52
53-55
56-58
59-61
62-64
65-67
68-70
71-73
74-76
77-79
>79
THI class

Figure 1

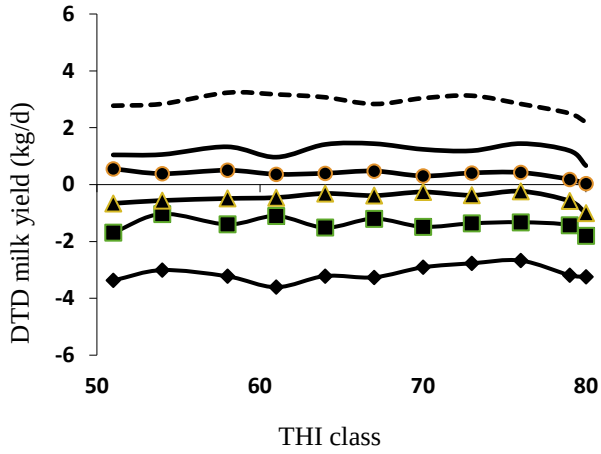


Figure 2a.

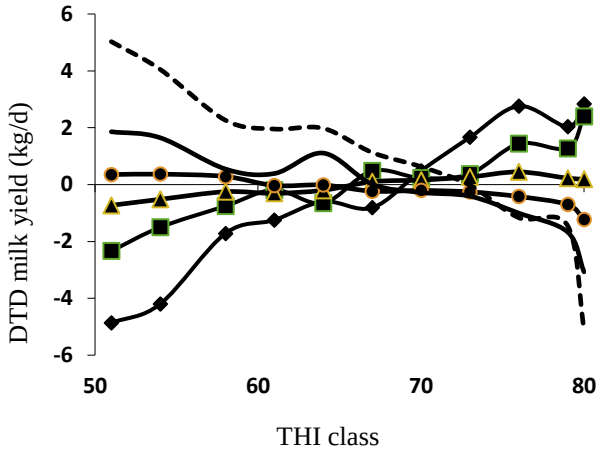
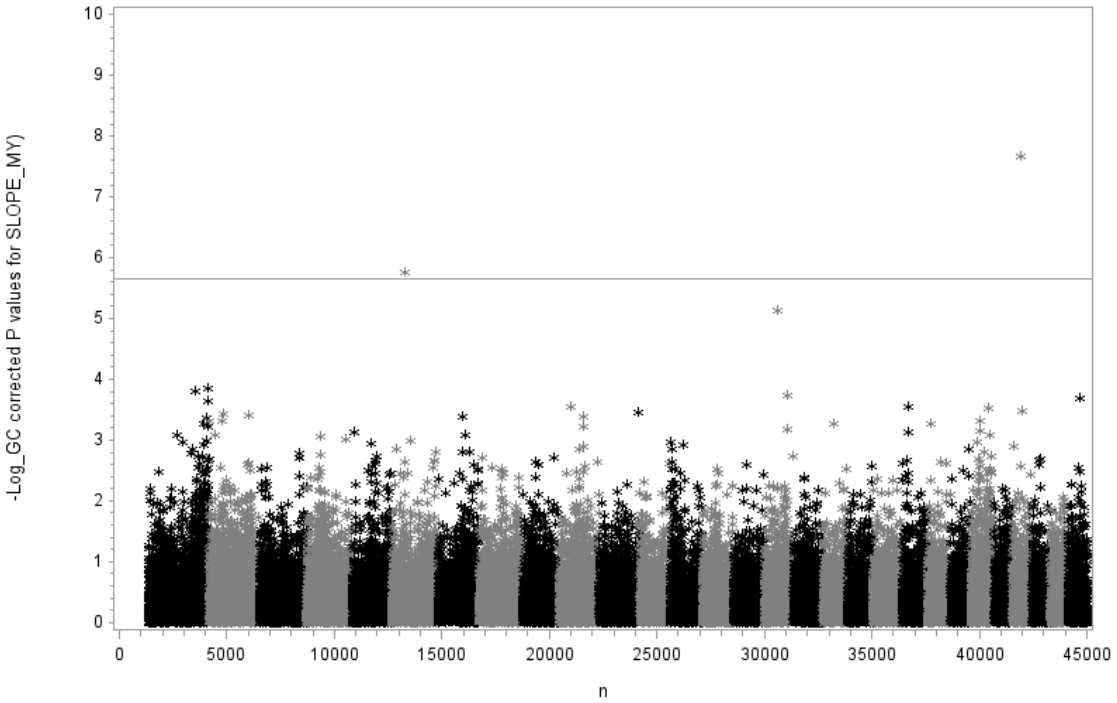
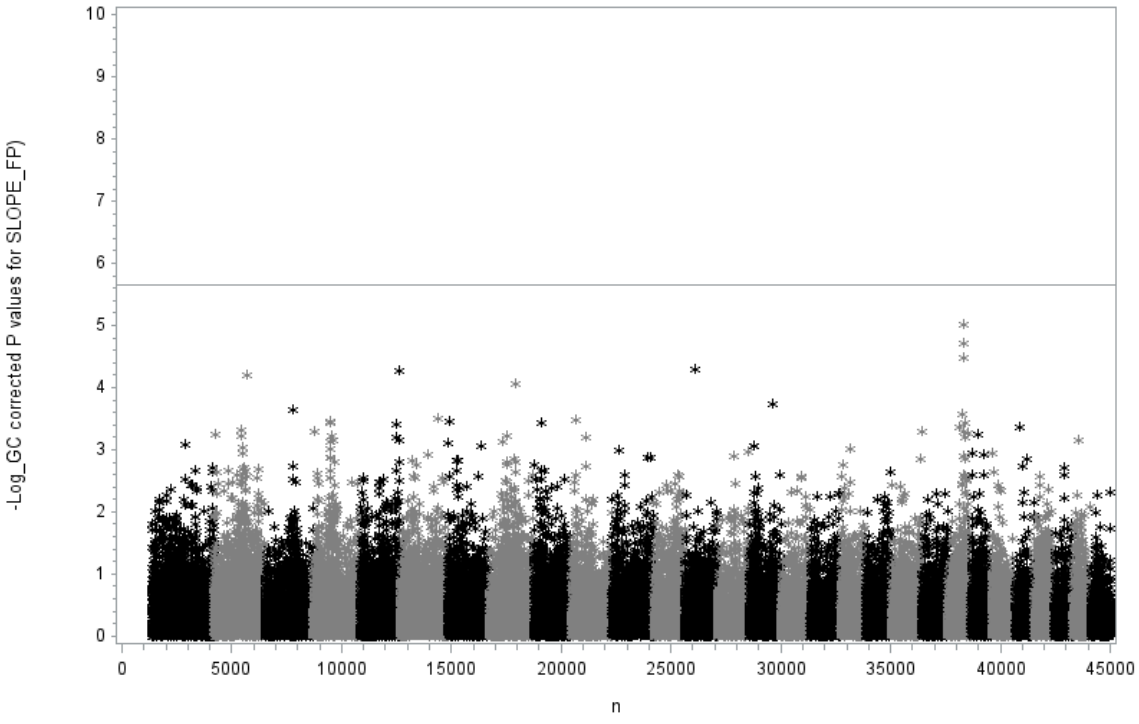


Figure 2b.

Figures 3


-Log_GC corrected P values for SLOPE_MY)
10
9
8
7
6
5
4
3
2
1
0
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
n


-Log_GC corrected P values for SLOPE_FP)
10
9
8
7
6
5
4
3
2
1
0
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
n

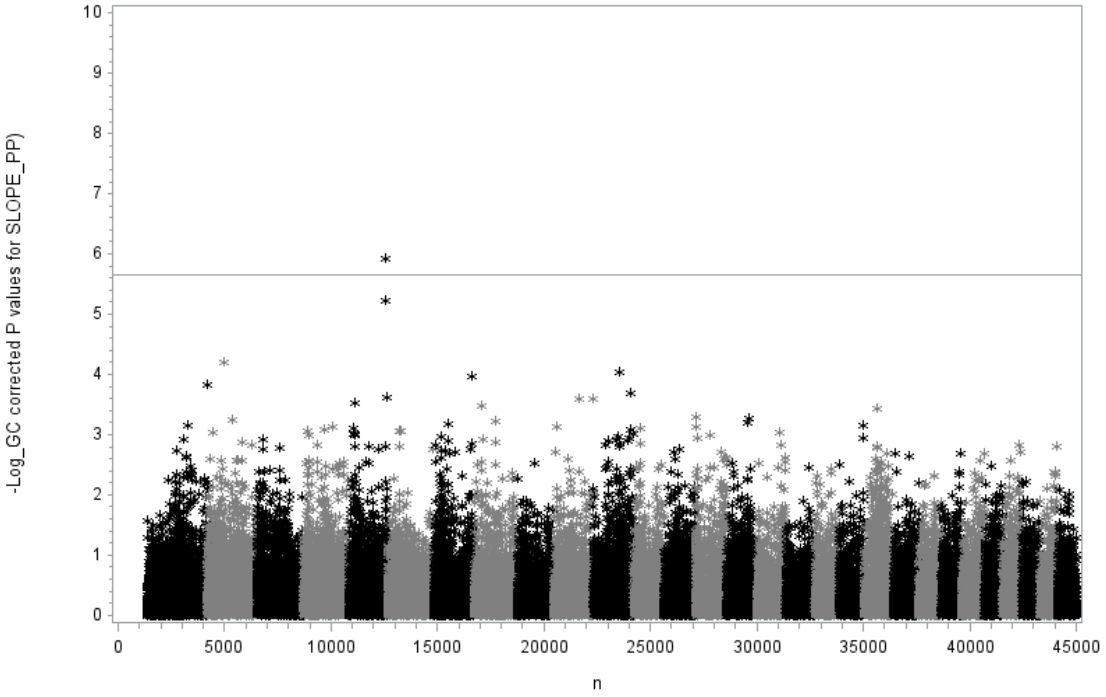
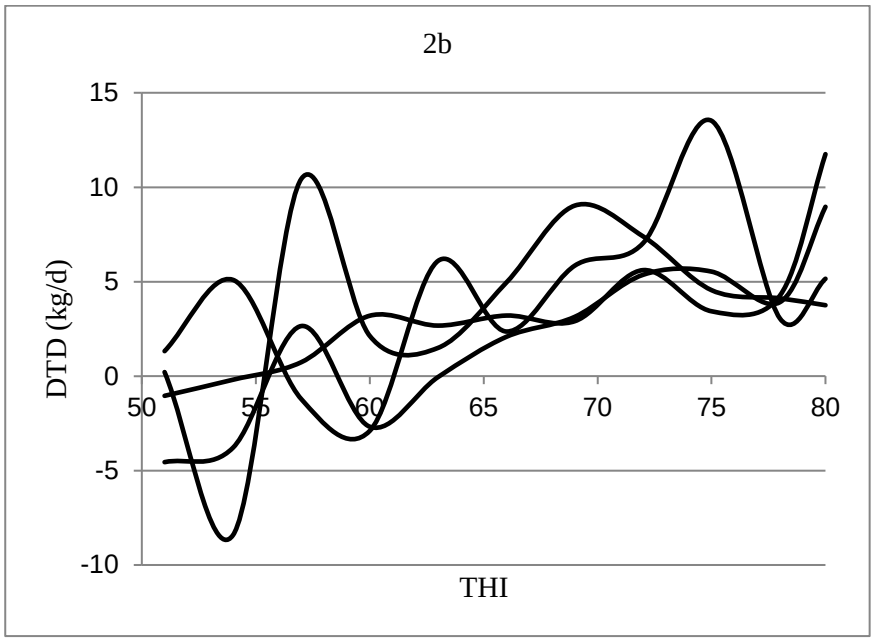

-Log_GC corrected P values for SLOPE_PP)
10
9
8
7
6
5
4
3
2
1
0
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
n

Figure 4.



Captions of figures

Figure 1. Distribution of test day records across different classes of Temperature humidity index (THI).

Figure 2a. Average curves of daughter trait deviations for milk yield of groups of bulls of different LEVEL 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 curves of daughter trait deviations for milk yield of groups of bulls of different SLOPE 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.

Figures 3. Genome-wide association study of the scores of the second principal component (SLOPE) for milk yield (figure 3a), fat percentage (figure 3b), and protein percentage (figure 3c). SLOPE_MY). On the vertical axis is reported the negative logarithm of the P value corrected for the stratification of the population. On the horizontal axis are reported the SNP ordered by their position and by chromosome. The dashed line corresponds to a FDR of 0.10.

Figure 4. Individual patterns of daughter trait deviations for milk yield ((DTD_MY) that have scores >2 and <-2 for the LEVEL and SLOPE components respectively.