Animal (2019), **13:7**, pp 1350–1357 © The Animal Consortium 2018
doi:10.1017/S1751731118003191

animal

Genome-wide association study for longevity in the Holstein cattle population

R. Steri[1], B. Moioli[1†], G. Catillo[1], A. Galli[2,3] and L. Buttazzoni[1]

[1]*Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), via Salaria 31, 00015 Monterotondo, Italy;* [2]*Istituto Sperimentale Italiano Lazzaro Spallanzani, Loc. La Quercia, 26027 Rivolta d'Adda, Italy;* [3]*Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), viale Piacenza 29, 26900 Lodi, Italy*

(Received 24 October 2017; Accepted 26 October 2018; First published online 3 December 2018)

Longevity is one of the most important traits determining dairy cow profitability. In the last decades dairy cows suffered a lowering in the age at culling. With the aim to identify the genes involved in longevity, dates of birth, yields, dates of calving during lifespan and culling dates were collected for 946 culled cows which had been genotyped with the Bovine High Density panel. Using the GenABEL package in R, genome-wide association analysis was performed on three potential traits of longevity: (1) 'days in production,' (2) 'days in herd,' (3) number of calvings over lifespan.' Five genome-wide significant single nucleotide polymorphisms (SNPs) associated with all three longevity traits were detected. Several consecutive SNPs identified on chromosomes 16 and 30 indicated the presence of two suggestive quantitative trait loci (QTL). The genes comprised in the QTL regions had biological functions related to fertility, reproductive disorders, heat stress and welfare of cows. These findings might contribute to improving breeding strategies to improve longevity.

Keywords: dairy cows, longevity, resilience, genome-wide study, fertility

Implications

The increase in milk yield has been accompanied by declining fertility, increasing metabolic problems and reduced longevity. Selection for long-living cows is hampered by the generally low heritability of functional traits, but due to the large amount of genetic variance in most dairy cattle populations great opportunities exist for selection. Genome-wide association studies (GWAS) are a powerful tool for unraveling the genetic background of complex traits. The availability of high-density marker panels, consisting of almost 800 thousand polymorphic markers, together with data recording of many cows under selection, allow for the use of estimation methods which might contribute to remodel selection schemes to be adopted by breeding organizations in their goal to improve longevity and welfare.

Introduction

Longevity is one of the most important traits determining dairy cow profitability: improved longevity results in increased proportion of more productive mature cows and reduced cost of replacement (Vollema and Groen, 1996; Sewalem *et al*., 2008; Morek-Kopec and Zarnecki, 2012). However, Oltenacu and Broom (2010) called attention to the increase in milk yield in the last 40 years for its association with declining longevity, this association being endorsed by Abfalter *et al*. (2016) who compared breeding values of cows with exceptional longevity with their contemporary herdmates (61 988 German Holstein cows): exceptional cows had on average significantly lower breeding value for milk, fat and protein yield but significantly higher breeding value for somatic cell score and functional longevity.

On the other hand, Pinedo *et al*. (2014) underlined that the main cause of culling of Holstein cows is low production. The risk of being culled in first lactation for low milk yield was confirmed by a recent longevity study on 12 045 Polish Holstein cows in 1371 herds (Sawa and Bogucki, 2017); these authors pointed out that age at culling in high producing herds (>8000 kg) was less than 3.93 years for cows yielding less than 7000 kg in first lactation, while age at culling ranged from 5.33 to 5.46 years for cows yielding more than 7000 kg in first lactation.

De Vries *et al*. (2010) reported that in the United States, during the years 2001–2006, 37% of cows in large herds were culled in their first parity and 83% within the first three parities.

† E-mail: bianca.moioli@crea.gov.it

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press

As both direct longevity (actually observed length of productive life) and functional longevity (adjusted for production) are only known when cows have completed their life, survival analysis, which combines records of culled and living animals (Ducrocq, 2005) was proposed to obtain earlier indicators of longevity. It is known that some traits are significantly related to longevity and that conformation traits and fertility represent reliable predictors of survival (Vukasinovic *et al*., 2001). In recent years, direct and indirect selection for calving ease, fertility, longevity, feet and legs, and indirect health traits have become common, and there has been an increase in genetic trends of functional traits (Egger-Danner *et al*., 2015).

Selection for long-living cows enjoyed considerable scientific and commercial interest but few GWAS were yet performed on longevity: in a recent study, Nayeri *et al*. (2017) targeted daughter survival of the proof bulls from the North American Holstein population, while Zhang *et al*. (2016), used the longevity breeding value of progeny-tested bulls from three Nordic cattle breeds of Denmark, Finland and Sweden. Aim of the present work was therefore to use data on direct longevity from over 900 dairy cows to identify candidate genes for longevity, through a GWAS performed with the high-density single nucleotide polymorphisms (SNPs) panel.

Material and methods

Animals and traits

The used data set, composed by one thousand herdbook registered Holstein cows from 302 herds of the most specialized dairy farming area of Northern Italy (Lombardia), was chosen for performing a GWAS on direct longevity, so to have the highest variability in this trait. The cows were in fact chosen in the two tails of the distribution of a longevity index calculated by the Italian National Holstein Breeders Association (ANAFI). Details of this index, estimated by a sire-maternal grandsire model which includes milk yield, somatic cell scores and scores for udder, legs and feet, weighed by the genetic correlations between trait and longevity, are given in the Supplementary Table S1. However, selection in the Italian Holstein is performed based on a productive index (productivity, functional traits, morphology) and the longevity index is provided to the farmers merely to help decisions on the choice of the replacements.

All cows were genotyped with the Bovine High Density (777 K) BeadChip (Illumina Inc., San Diego, CA, USA). As the aim of the present work was to investigate the genomics of direct longevity, the following traits were defined and analyzed: (1) 'days in production,' expressing the days from first calving to culling, which is the most commonly used longevity measure (Zavadilova and Štípková, 2012); (2) 'days in herd,' that is, the days from birth to culling, which also takes into account growth and health of replacements (Heinrichs and Heinrichs, 2010); (3) 'calvings,' that is, the number of calvings of a cow during lifespan, as true expression of productivity according to Vollema and Groen (1996). All experimental females calved at least once, therefore in this study 'direct longevity' actually is 'direct longevity conditional on first calving.' The parameter 'days in herd' was then considered only for the cows which had already calved at least once.

Dates of birth, dates of all calvings during lifespan and culling date were collected and provided by the Italian Breeders Association (AIA), while genotypes were provided by Istituto Sperimentale Italiano Lazzaro Spallanzani (Rivolta d'Adda, Italy). DNA of all cows was sent for genotyping in a single batch; genotyped cows which had already been culled before the beginning of the present work were 946, belonging to 297 farms, born from 1999 to 2009, calving from 1 to 12 times in lifespan and had the following mean values for the traits to be analyzed: days in production: 1578.88 ± 665.07; days in herd: 2413.80 ± 670.32; calvings: 3.77 ± 1.65.

Genome-wide association

Pruning for Linkage Disequilibrium (LD) was performed as recommended by Price *et al*. (2008) who demonstrated that strong LD at a particular locus has a confounding effect on the global genome-wide variation and might reflect genetic variation only at that specific locus. Pruning was performed with PLINK (Purcell *et al*., 2007) using a variance genomic inflation factor = 2, a window of 50 SNPs and shifting the windows by 5 SNPs at each run. Furthermore, markers with missing data (⩾10%) and minor allele frequency ⩽1% were also excluded using PLINK. As all cows were successfully genotyped at >80% of the SNP, no further filtering on the cows was applied.

As a consequence of the selective genotyping, half of the herds in the data set had only one genotyped cow, and several herds had only two genotyped cows. As it was not possible to consider herd and year of calving in the model, in order to take into account herd management the following two factors were considered: (1) average daily milk yield of each herd/year classified in the following five classes: < 26 kg; 27 to 29 kg; 30 to 31 kg; 32 to 35 kg; >35 kg. (2) Average number of cows of each herd/year classified in the following four classes: <60; 61 to 130; 131 to 200; >200.

Genome-wide association analysis was performed with the GenABEL package in R (Aulchenko *et al*., 2007) which performs a simple linear regression marker-phenotype analysis. Phenotypes were trait deviations from the overall mean values of the raw phenotypes: days in production, days in herd and calvings. The following model was used:

$$\mathbf{Y}_i = \mathbf{X}_i\beta_i + \mathbf{w}_i h_i + \mathbf{Z}_i a_i + \mathbf{e}_i$$

where $\mathbf{Y}_i$ is the observation vector for the *i*th trait: days in production, days in herd and calvings; β_i the vector of fixed effects for trait *i*th (herd/year class of milk yield of each cow at time of her first calving: five levels; herd/year class of herd size of each cow at time of her first calving: four levels), $\mathbf{X}_i$ the matrix of incidence of $n \times p$ order

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press

associating each observation (n) to the pertinent level of fixed effects (p); h_i the effect size of marker, $\boldsymbol{w}_i$ a n by 1 vector of marker genotypes; a_i the vector of random polygenic effects $(\boldsymbol{a}_i \sim \boldsymbol{N}(0, \boldsymbol{G}\sigma_a^2))$, where σ_a^2 is the estimated polygenic variance component, and $\boldsymbol{G}$ the $n \times n$ genomic relationship matrix, $\mathbf{Z}_i$ the n x n matrix of incidence associating each animal; $\boldsymbol{e}_i$ a n by 1 vector of errors $\boldsymbol{e}_i \sim N(0, \boldsymbol{I}\sigma_e^2)$ where σ_a^2 is the environmental variance component, and I the $n \times n$ identity matrix.

Regression was performed using the 'egscore' function of the GenABEL package for taking into account the population substructure likely induced by selective genotyping. The 'egscore' function is based on the EIGENSTRAT method (Price *et al*., 2006) and accounts for population stratification by adjusting for the principal components of variation of genomic kinship matrix, using as default the first ten components. The Genomic Control (GC) approach was also used to correct for the inflated null distribution of the 'egscore' test statistics, using the inflation factor λ which is the observed median value of the χ^2 statistic for the null markers divided by the expected median value of the χ^2 statistic (Hinrichs *et al*., 2009). *P*-values of each marker effect after GC are directly produced as results of the analysis by the 'egscore' function of the GenABEL package.

Finally, to control the increased type I error rising from the multiple hypotheses simultaneously tested in the above model, the *P*-values of the effects of each SNP corrected for GC were further corrected using the Proc Multtest of SAS software (SAS Institute Inc., 2007), by applying the false discovery rate (FDR) test (Benjamini and Hochberg, 1995) which controls also the probability of erroneously rejecting even one of the true null hypotheses. Only the markers reaching a threshold level < 0.05 after FDR test were taken into consideration as signals of longevity.

The variance associated to the significant SNPs was estimated using the formula $V = 2pqa^2$ (Falconer and Mackay, 1996), where p and q are the frequencies of the minor and major alleles and a is the allelic substitution effect. This value was then divided by the total genomic variance of the model provided by the GenABEL package.

To verify whether significant SNPs, located on the same chromosome, encompassed the same quantitative trait loci (QTL), measure of LD between them was calculated using the Proc Allele of SAS software (SAS Institute Inc., 2007).

As Hinrichs *et al*. (2009) warned that the GC method has the disadvantage of unintentionally including true positives in the null set, so biasing association tests conservatively, the SNPs pruned out because of their LD with a significant one were retrieved from the PLINK file and reconsidered when falling in a region 500 Kbp upstream or downstream the significant one (Wu *et al*., 2013). This region was assumed to potentially encode a QTL for longevity.

Gene network analysis

The genomic regions encoding QTLs were explored and the annotated genes within these regions were retrieved from the NCBI database. Location and gene annotation were based on the UMD 3.1 genome assembly (https://www.ncbi.nlm.nih.gov/; annotation release 104).

Functional annotation tool of PANTHER classification System (Version10.0, Mi *et al*., 2013) was used to inspect Gene Ontology (GO) terms mapping to protein-coding genes and assign the genes to the three main categories, Biological Process (BP), which includes broad biological goals; Molecular Function (MF), which describes the tasks performed by individual gene products and Cellular Component (CC), which involves subcellular structures, locations and macromolecular complexes.

Results

The first finding of this study was that phenotypic values for direct longevity of cows born from 1999 to 2009 registered a continuous decrease (Table 1), while 305 days milk yield in first lactation increased till 2007 then remained constant (Table 1). This suggests that, beyond first lactation yield, other events have contributed to culling.

Genome-wide association analysis

The working data set after pruning for LD and quality control of the genotypes obtained with the SNP panel (777 962 SNPs) contained 946 samples and 135 992 SNPs.

Table 1 *Average parameters of true longevity and 305 days first lactation by year of birth of the analyzed cows*

		days_production		days_herd		calvings		Milk 305d (kg)	
year_birth	*N* cows	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1999 to 2001	10	3964.3	495.1	4907.6	515.4	9.4	1.6	8115.4	1923.8
2002	6	3235.7	413.7	4037.5	397.5	7.6	1.4	8514.3	1564.1
2003	15	2726.1	321.8	3612.3	321.5	6.2	1.1	9086.5	1150.6
2004	48	2361.3	453.6	3223.2	439.8	5.5	1.1	9101.3	1589.8
2005	93	2181.8	444.9	3020.7	426.1	5.1	1.3	9100.3	1589.7
2006	151	1790.3	470.9	2625.2	455.0	4.3	1.4	9350.4	1710.8
2007	242	1450.5	491.1	2304.6	466.1	3.5	1.3	9466.9	1510.6
2008	289	1228.1	480.0	2058.8	468.0	2.9	1.3	9421.8	1639.4
2009	92	1090.4	460.3	1852.7	462.4	2.7	1.1	9334.91	1619.1

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press

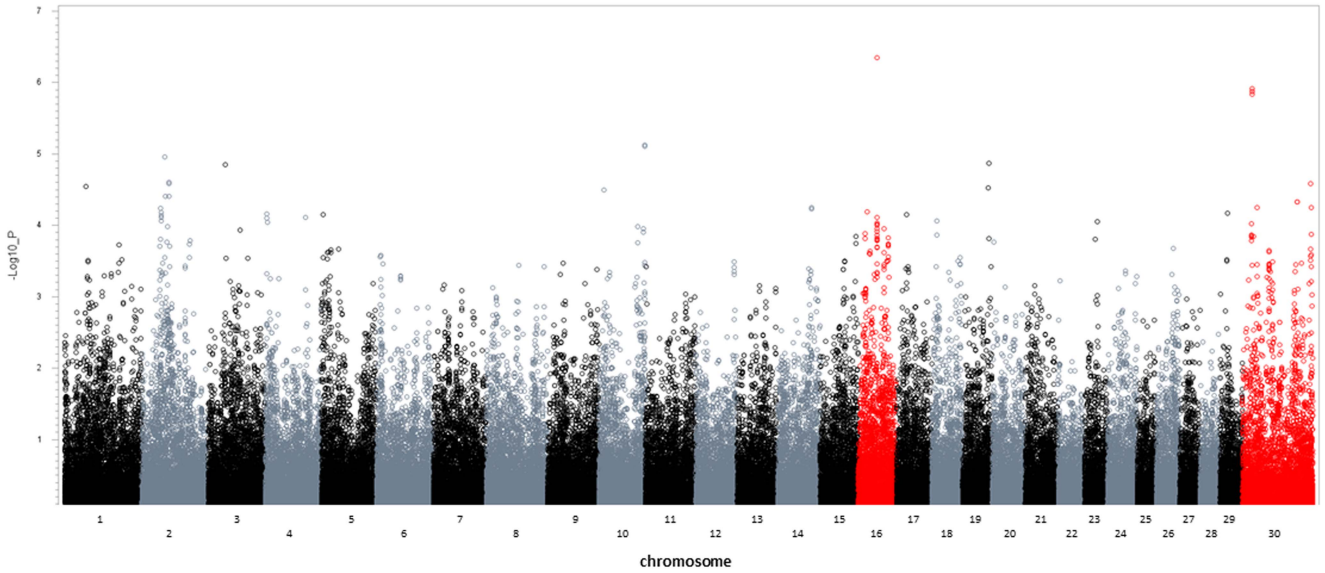


Figure 1 Manhattan plot displaying the results of the genome wide association with respect to the genomic position for 'days in production', in the analyzed cows. On the Y axis, − log10 of raw *P*-value of each single nucleotide polymorphism (SNPs) are reported. Significant SNPs after genomic control and False Discovery Rate corrected have − log10 *P*-value > 5.84. Alternate black and gray color were used to distinguish chromosomes 1 to 30. The chromosomes 16 and 30 in red color are those presenting significant SNPs.

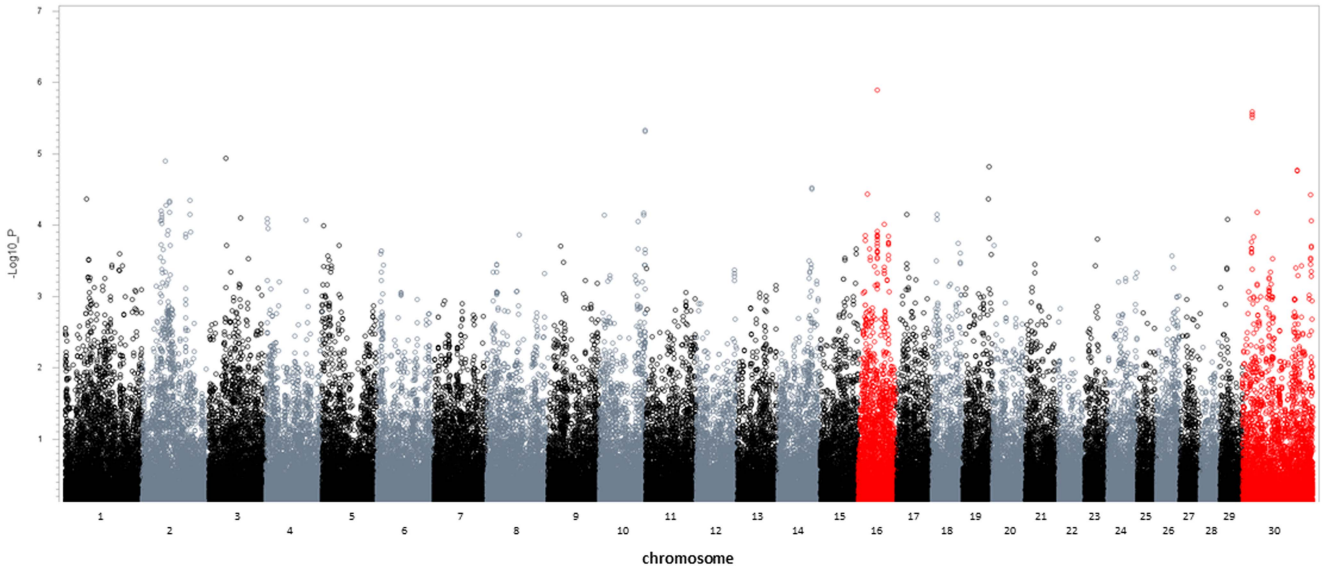


Figure 2 Manhattan plot displaying the results of the genome wide association with respect to the genomic position for 'days in herd', in the analyzed cows. On the Y axis, − log10 of raw *P*-value of each single nucleotide polymorphism (SNPs) are reported. Significant SNPs after genomic control and False Discovery Rate corrected have − log10 *P*-value > 5.84. Alternate black and gray color were used to distinguish chromosomes 1 to 30. The chromosomes 16 and 30 in red color are those presenting significant SNPs.

Genome-wide Manhattan plots displaying the GWAS results (− log10 of *P*-value of each SNP) with respect to their genomic position were shown in Figures 1 to 3 for the three analyzed traits. The five most significant SNPs – one on Bos taurus autosome (BTA) 16, four on BTA 30 – were the same for all three traits, though ranking differently, and were reported in Table 2. In the table, beyond the name of the SNP, the chromosome and the position on the assembly UMD v3.1, the allelic substitution effect with corresponding χ^2 values were reported for each of the three traits. Raw *P*-values, as well as *P*-values corrected for GC and for FDR were also reported in Table 2. The five SNPs showed significant effects (FDR-corrected *P*-value <0.05) on 'days in production' and 'calvings' while they only tended to significance (FDR-corrected *P*-value < 0.08) for 'days in herd,' this likely reflecting the higher variability of the latter trait which is also affected by the age at first calving. Average allelic substitution effects (absolute values) for the significant SNPs were 105.34 ± 1.88 and 102.49 ± 1.11 days for 'days in production' and 'days in herd,' respectively; and 0.26 ± 0.01 for 'calvings.' Minor allele frequency was 0.11 for SNPs on both BTA 16 and 30. Linkage disequilibrium between each

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press

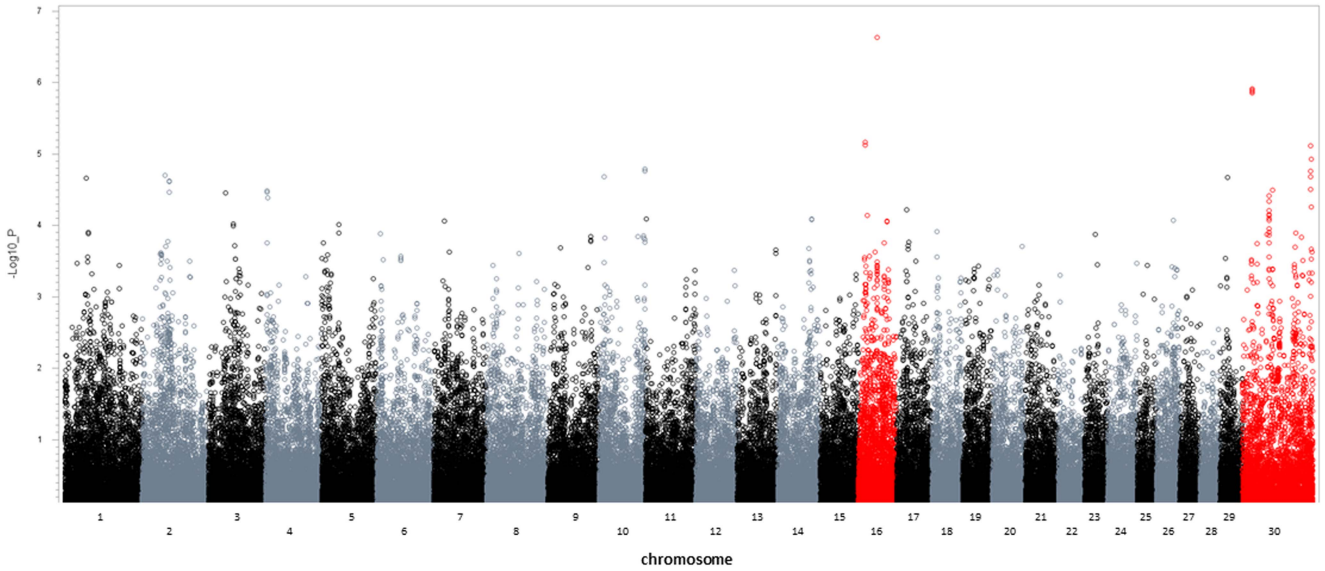


Figure 3 Manhattan plot displaying the results of the genome wide association with respect to the genomic position for 'calvings', in the analyzed cows. On the Y axis, −log10 of raw *P*-value of each single nucleotide polymorphism (SNPs) are reported. Significant SNPs after genomic control and False Discovery Rate corrected have −log10 *P*-value >5.84. Alternate black and gray color were used to distinguish chromosomes 1 to 30. The chromosomes 16 and 30 in red color are those presenting significant SNPs.

Table 2 *Significant single nucleotide polymorphisms for the three longevity traits, found in the analyzed cows*

	Days in production					Days in herds					Calvings				
SNP[1]	B[2]	χ^2	Raw *P*-value	GC *P*-value[3]	FDR[4]	B[2]	χ^2	Raw *P*-value	GC *P*-value[3]	FDR[4]	B[2]	χ^2	Raw *P*-value	GC *P*-value[3]	FDR[4]
BovineHD1600012095 BTA 16: 43250880	108.69	63.90	1.4E-15	4.4E-7	0.03	104.06	58.04	1.9E-14	1.2E-6	0.08	0.28	66.90	2.8E-16	2.3E-7	0.03
BovineHD3000005842 BTA 30: 17567524	104.34	58.01	2.6E-14	1.4E-6	0.03	101.45	54.42	1.6E-13	2.5E-6	0.08	0.26	58.25	2.3E-14	1.4E-6	0.03
BovineHD3000005849 BTA 30: 17588787	104.34	58.43	2.1E-14	1.3E-6	0.03	101.45	54.45	1.6E-13	2.8E-6	0.08	0.26	58.39	2.3E-14	1.4E-6	0.03
BovineHD3000005850 BTA 30: 17591386	104.48	58.41	2.1E-14	1.3E-6	0.03	101.59	54.81	1.3E-13	3.0E-6	0.08	0.26	58.67	1.9E-14	1.3E-6	0.03
BovineHD3000005851 BTA 30: 17595118	104.87	58.88	1.7E-14	1.2E-6	0.03	101.99	55.25	1.1E-13	2.5E-6	0.08	0.26	58.87	1.7E-14	1.3E-6	0.03

For each effect, Raw *P*-value, *P*-value corrected after genomic control, and False Discovery Rate (FDR) *P*-value are reported.
[1]Single nucleotide polymorphism name (SNP), Bos taurus autosome (BTA) and position on BTA.
[2]Allelic substitution effect.
[3]*P*-value corrected after genomic control (GC).
[4]*P*-value corrected for FDR.

pair of the four SNPs of chromosome 30 varied between 99.3% and 100%; moreover, since the effect of the four SNPs as well as the values of χ^2 were very similar (Table 2), they should be considered a unique QTL. Therefore, the variance explained by one marker of each QTL, calculated as in Falconer and Mackay (1996) and divided by the total genomic variance of the model, was 9.3% and 8.5% for days in production, 8.5% and 8.1% for days in herd and 10% and 8.7% for calvings, for BovineHD1600012095 and BovineHD3000005850, respectively.

The pruned out SNPs located in a genomic region spanning 500 Kbp upstream or downstream the significant SNPs were reconsidered and were reported in the Supplementary Table S2, with indication of their position on the chromosome. In that Table, the five significant SNPs of Table 2 were evidenced in bold. The two genomic regions on BTA 16 and 30 encompassing the significant SNPs of Table 2, as well as the SNPs in LD with them of Table S2, encoded ten genes (Table 3) which were analyzed through PANTHER's functional annotation tool to inspect GO terms mapping to protein-coding genes in Bos Taurus. Genes were classified in the GO categories BP, MF and CC (Table 4). Under the BP classification, the detected genes are involved in cellular processes, metabolic process, response to stimuli and immune system process. Under the MF classification (Table 4), the detected genes are involved in binding and

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press

Table 3 *Summary of the most interesting candidate genes for longevity traits found in the analyzed cows*

	Position of significant SNPs		Gene position			
BTA	Start	End	Start	End	Gene symbol	Gene name
16	43250880	43697344	43256416	43268600	*UBIAD1*	UbiA prenyltransferase domain containing 1
			43273563	43397006	*MTOR*	Mechanistic target of rapamycin
			43328466	43334550	*ANGPTL7*	Angiopoietin like 7
			43403612	43427085	*EXOSC10*	Exosome component 10
			43445482	43449515	*SRM*	Spermidine synthase
			43463043	43478362	*MASP2*	Mannan-binding lectin serine peptidase 2
			43479938	43486540	*TARDBP*	TAR DNA-binding protein
			43686458	43832154	*CASZ1*	Castor zinc finger 1
30	17567524	18062626	17305528	17770802	GPC3	Glypican 3
			18048826	18085910	*PHF6*	PHD finger protein 6

BTA = Bos taurus autosome; SNP = single nucleotide polymorphism name.

catalytic activity; under the CC classification genes belonged to cell part, organelle and extra cellular region categories.

Discussion

While GWAS for economic traits – fertility, mastitis susceptibility, production – in dairy cows has been enjoying considerable scientific interest in the past 10 years (Pryce *et al*., 2010; Wu *et al*., 2013), only recently longevity has been targeted in GWAS, using either the 50 K SNP panel (Zhang *et al*., 2016) or the 777 K SNP panel (Nayeri *et al*., 2017). Both authors used longevity breeding values of progeny-tested bulls as phenotype for the GWAS, while the present study analyzed direct longevity data of cows from the two tails of the distribution of a composite index made of proxy traits for longevity.

As the use of the high-density marker panel (777 K) might increase the confounding effect of LD on the global genome-wide variation, also considering that the sampled cows were likely affected by sub-population stratification, pruning for LD was considered an appropriate solution. The working data set was made by 135 992 SNPs, and the percentage of pruned out SNPs was similar to other GWAS performed on humans (Tassi *et al*., 2015), horses (Petersen *et al*., 2013) and pigs (Lopes *et al*., 2013).

Each of the two QTLs on chromosomes 16 and 30, common to the three longevity traits, explains between 8% and 10% of the genomic variability of each trait; however, de los Campos *et al*. (2015) warn against the use of marker-based regressions for inferences about population parameters such as the genomic variance; these authors showed that the genomic heritability and the trait heritability parameters are equal only when all causal variants are typed. Using both simulated and real data Yang *et al*. (2010) have used whole-genome regression models with the purpose of obtaining inferences of genomic parameter that resemble more closely to those based on pedigrees, concluding that the different scenarios, involving distantly or more closely related individuals, as well as the LD between the markers and the QTLs highly affect the estimation. Speed *et al*. (2012) emphasized that contributions of markers to h^2 are overestimated from causal variants in regions of high LD and are underestimated in regions of low LD. By acknowledging such warnings, the five significant SNPs will be taken simply as an indication of two QTLs associated to longevity, and the genes encoded by the QTLs will be examined under the assumption that they might have a role in influencing longevity in cows.

Table 4 *Gene Ontology (GO) classification of putative longevity genes found in the analyzed cows*

Class ontology	Genes
Biological Process	
Cellular process (GO:0009987)	*MTOR*
Metabolic process (GO:0008152)	*UBIAD1 MTOR EXOSC10 MASP2 PHF6*
Response to stimulus (GO:0050896)	*MTOR MASP2*
Immune system process (GO:0002376)	*MASP2*
Molecular Function	
Binding (GO:0005488)	*EXOSC10 ANGPTL7 PHF6*
Catalytic activity (GO:0003824)	*UBIAD1 MTOR EXOSC10 MASP2*
Cellular Component	
Cell part (GO:0044464)	*MTOR*
Organelle (GO:0043226)	*ANGPTL7*
Extracellular region (GO:0005576)	*MTOR*

Moreover, the five significant SNPs were in LD with a number of consecutive SNPs (Supplementary Table S2) which further support the existence of two QTLs affecting longevity on BTA 16 and 30.

The GO (Table 4) was helpful in ascribing the contribution of the 10 detected genes (Table 3) encoded by the QTLs to different functional pathways, so to elucidate the potential role they might play in longevity. Under the BP classification, the identified categories – cellular processes, metabolic process, response to stimuli, immune system process – were also the most represented categories within the highly scored target genes of up-regulated and down-regulated

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press

microRNAs in cows with metritis and normal uterus (Kasimanickam and Kastelic, 2016), corroborating the connections between longevity, fertility and reproductive disorders. Moreover, metabolic process, and response to stimulus were also the most abundant sub-categories of the differentially expressed genes in response to heat stress in mammary epithelial cells of river buffalo (Kapila *et al.*, 2016), suggesting further connections between stress, welfare and longevity.

Under the MF classification, the detected genes were involved in binding and catalytic activity, and these sub-categories were the most represented by the copy number variation (CNV) genes identified by Keel *et al.* (2016) who explored exome sequence from 175 influential sires representing 10 breeds, with the aim to investigate evolutionary mechanisms of variation in the bovine genome. Interestingly, of the 110 CNV genes identified by these authors 35% were involved in binding and 23% in catalytic activity. Furthermore, CNVs are prone to be subjected to selective constraints (Keel *et al.*, 2016), therefore genes influencing longevity are likely to be more susceptible to selective pressure and are presumably affected by the strong selection pressure on productive traits in dairy cows.

The detected genes under the CC classification belonged to cell part, organelle and extra cellular region categories, which were the most represented categories by the differentially expressed genes in ovaries of two local Chinese sheep breeds, characterized by high-fecundity and low-fecundity, respectively, suggesting a potential role of the genes of this category in influencing fertility (Chen *et al.*, 2014).

The significant marker of chromosome 16 identified in the present study – BovineHD1600012095 and the immediately downstream following marker – BovineHD1600012099 – located at position 43250880 and 43267903, respectively (Table 3), encompass the *UBIAD1* gene (Table 3), which is a K2 vitamin biosynthetic enzyme. Performing a GWAS by the 50 K SNP Chip on fertility traits and reproductive disorders on Simmental cows in Austria, Rauter (2014) identified a marker on chromosome 16: 43.33 Mbp significantly associated to stillbirth. Therefore, Rauter (2014) suggested that *UBIAD1* gene has a role in causing stillbirth. It is remarkable that mice lacking of this gene die as embryos (Nakagawa *et al.*, 2014). Moreover, the QTL of chromosome 16, consisting of eight consecutive SNPs in LD (Table 3), encompasses also the *MTOR* gene, implicated in the regulation of both the innate and adaptive immune responses (Soliman, 2013).

Conclusion

To our knowledge, this is the first GWAS performed on three indicators of direct longevity. GWAS was performed on a sample of Holstein cows showing the broadest variability in longevity; the high-density marker panel, 14 times more dense than the commonly used 50 K panel, was used and novel regions affecting longevity were observed. Some of the putative genes for longevity identified in this study have been already associated to fertility in cattle, while most of them play a role in basic cellular processes also described, through the GO analysis categories, in gene expression studies focusing on fertility, reproductive disorders, heat stress and welfare. The present study therefore, through genomic data, might suggest that culling is associated to reduced reproductive and productive efficiency and resilience of cows.

Acknowledgements

This study was supported by the research program 'WELL4DOP' funded by the Italian Ministry of Agriculture (D.M. 6126/7303/15 of 20 March 2015). Data were provided by the research program 'RIPROWEL' (Ministry of Agriculture DM 13576 of 24/05/2011). The authors thank Graziella Bongioni and Caterina Cambuli of the Istituto Sperimentale Italiano Lazzaro Spallanzani (ISILS, Rivolta d'Adda, Italy), and the Italian Holstein Breeders Association (ANAFI, Cremona, Italy) and the AIA (Roma, Italy) for providing genetic data and phenotypes.

Declaration of interest

The authors declare no conflicts of interest

Ethics statement

No new ethical approval was required since the data for this study were provided by the research program 'RIPROWEL' (Ministry of Agriculture DM 13576 of 24/05/2011).

Software and data repository resources

None of the data were deposited in an official repository.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731118003191

References

Abfalter K, Brade W and Distl O 2016. Comparison of breeding values among cows withexceptional longevity and their contemporary herdmates in German Holsteins. Archives Animal Breeding 59, 71–77.

Aulchenko YS, Ripke S, Isaacs A and van Duijn CM 2007. GenABEL: an R package for genome-wide association analysis. Bioinformatics 23, 1294–1296.

Benjamini Y and Hochberg Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57, 289–300.

Chen HY, Shen H, Jia B, Zhang YS, Wang X and Zeng X 2014. Differential gene expression in ovaries of qira black sheep and hetian sheep using RNA-seq technique. PLoS One 10, e0120170.

de los Campos G, Sorensen D and Gianola D 2015. Genomic heritability: what is it? PLoS Genetics 11, e1005048.

De Vries A, Olson JD and Pinedo PJ 2010. Reproductive risk factors for culling and productive life in large dairy herds in the eastern United States between 2001 and 2006. Journal of Dairy Science 93, 613–623.

Ducrocq V 2005. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. Animal Science 80, 249–256.

Egger-Danner C, Cole JB, Pryce JE, Gengler N, Heringstad B, Bradley A and Stock AF 2015. Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal 9, 191–207.

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press

Falconer D and Mackay T 1996. Introduction to Quantitative Genetics. Pearson Education Limited, Harlow, Essex.

Heinrichs AJ and Heinrichs BS 2010. A prospective study of calf factors affecting first-lactation and lifetime milk production and age of cows when removed from the herd. Journal of Dairy Science 94, 336–341.

Hinrichs AL, Larkin EK and Suarez BK 2009. Population stratification and patterns of linkage disequilibrium. Genetic Epidemiology 33 (suppl. 1), S88–S92.

Kapila N, Sharma A, Kishore A, Sodhi M, Tripathi PK, Mohanty AK and Mukesh M 2016. Impact of heat stress on cellular and transcriptional adaptation of mammary epithelial cells in riverine buffalo. PLoS One 11, e0157237.

Kasimanickam V and Kastelic J 2016. Circulating cell-free mature microRNAs and their target gene prediction in bovine metritis. Scientific Reports 6, 29509.

Keel BN, Lindholm-Perry AK and Snelling WM 2016. Evolutionary and functional features of copy number variation in the cattle genome. Frontiers in Genetics 7, 207.

Lopes MS, Silva FF, Harlizius B, Duijvesteijin N, Lopes PS, Guimaraes SE and Knol E 2013. Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. BMC Genetics 14, 92.

Mi H, Muruganujan A and Thomas PD 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Research 41, D377–D386.

Morek-Kopec M and Zarnecki A 2012. Relationship between conformation traits and longevity in Polish Holstein Friesian cattle. Livestock Science 149, 53–61.

Nakagawa K, Sawada N, Hirota Y, Uchino Y, Suhara Y, Hasegawa T, Amizuka N, Okamoto T, Tsugawa N, Kamao M, Funahashi N and Okano T 2014. Vitamin K2 biosynthetic enzyme, UBIAD1 is essential for embryonic development of mice. PloS One 9, e104078.

Nayeri S, Sargolzaei M, Abo-Ismail MK, Miller S, Schenkel F, Moore SS and Stothard P 2017. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. Journal of Dairy Science 100, 1246–1258.

Oltenacu PA and Broom DM 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. Animal Welfare 19 (S), 39–49.

Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns M, Borges AS, Brama P, da Câmara Machado A, Distl O, Felicetti M, Fox-Clipsham L, Graves KT, Guérin G, Haase B, Hasegawa T, Hemmann K, Hill MW, Leeb T, Lindgren G, Lohi H, Lopes MS, McGivney BA, Mikko S, Orr N, Penedo MC, Piercy RJ, Raekallio M, Rieder S, Røed KH, Silvestrelli M, Swinburne J, Tozaki T, Vaudin M, Wade CM and McCue ME 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS One 8, e54997–e54997.

Pinedo PJ, Daniels A, Shumaker J and De Vries A 2014. Dynamics of culling for Jersey, Holstein, and Jersey × Holstein crossbred cows in large multibreed dairy herds. Journal of Dairy Science 97, 2886–2895.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick MA and Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38, 904–909.

Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, Ge D, Rotter JI, Torres E, Taylor KD, Goldstein DB and Reich D 2008. Long-range LD can confound genome scans in admixed populations. The American Journal of Human Genetics 83, 127–147.

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ and Savin K 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. Journal of Dairy Science 93, 3331–3345.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–575.

Rauter M 2014. Genome-wide association study for fertility related traits in Austrian Fleckvieh cattle. Master Thesis, University of Natural Resources and Life Sciences, Vienna, Austria.

SAS Institute Inc. 2017. SAS/STAT User's Guide, Version 9.1. SAS Institute Inc., Cary, NC, USA.

Sawa A and Bogucki M 2017. Longevity of cows depending on their first lactation yield and herd production level. Annals of Animal Science 17, 1171–1183.

Sewalem AF, Miglior F, Kistemaker GL, Sullivan P and van Doormaal BJ 2008. Relationship between reproduction traits and functional longevity in Canadian Dairy Cattle. Journal of Dairy Science 91, 1660–1668.

Speed D, Hemani G, Johnson MR and Balding DJ 2012. Improved heritability estimation from genome-wide SNPs. American Journal of Human Genetics 91, 1011–1021.

Soliman G. 2013. The role of mechanistic target of rapamycin (mTOR) complexes signaling in the immune responses. Nutrients 5, 2231–2257.

Francesca T, Ghirotto S, Mezzavilla M, Vilaça ST, Santi LD and Barbujani G 2015. Early modern human dispersal from Africa: genomic evidence for multiple waves of migration. Investigative Genetics 6, 13.

Vollema AR and Groen AF 1996. Genetic parameters of longevity traits of an upgrading population of dairy cattle. Journal of Dairy Science 79, 2261–2267.

Vukasinovic N, Moll J and Casanova L. 2001. Implementation of a routine genetic evaluation for longevity based on survival analysis techniques in dairy cattle populations in Switzerland. Journal of Dairy Science 84, 2073–2080.

Wu Xiaoping FM, Liu L, Wang S, Liu J, Ding X, Zhang S, Zhang Q, Zhang Y, Qiao L, Lund MS, Su G and Sun D 2013. Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. BMC Genomics 14, 897.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME and Visscher PM 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42, 565–569.

Zavadilová L and Štípková M 2012. Genetic correlations between longevity and conformation traits in the Czech Holstein population. Czech Journal of Animal Science 57, 125–136.

Zhang Q, Guldbrandtsen B, Thomasen JR, Lund MS and Sahana G 2016. Genome-wide association study for longevity with whole-genome sequencing in 3 cattle breeds. Journal of Dairy Science 99, 7289–7298.

https://doi.org/10.1017/S1751731118003191 Published online by Cambridge University Press