Dipartimento di Scienze Medico-Veterinarie

Corso di laurea Magistrale a ciclo unico in Medicina Veterinaria

Genes related to lactation curve persistence in Italian Holstein cattle

Geni correlati alla persistenza della curva di lattazione nelle bovine di razza Holstein Italiana

Relatore:
Chiar.mo Prof. Alberto Sabbioni

Correlatore:
Dott.ssa Michela Ablondi

Laureando:
Andrea Zanotti

Anno accademico 2021-2022

Summary

Abstract

The persistence of the lactation curve is a very important parameter in dairy farming because of its effect on fertility, health, profitability of a farm. Furthermore, paying attention to it affects might change the management of the farm as well. Unfortunately, nowadays little is known about persistence, mainly due to a standard breeding system that relies on 305-day lactations and does not allow animals to express their persistence characteristics to the fullest. The purpose of this thesis thus, was to analyze the current level of persistence in two different herds of Holstein cattle managed under different conditions.

Test day records were collected for all the lactating animals present in the two farms in a two-year period 11/19 through 11/21, which allowed to model the lactation curve. In these two farms all animals were genotyped with several different genotype panels. The genomes were thus imputed to get 80K SNPs per each animal and therefore perform a Genome Wide Association Study (GWAS).

A total of about 12,993 test day records belonging to 1035 cows were collected.

The ANOVAs studies were done to compare the effect of farm, parity and season and the interaction between the above-mentioned effect. These studies were necessary to correct for the environmental effect the parameters obtained from the lactation curve: persistence, peak time and peak yield.

As a result, from ANOVAs, it was noted that the effect of season, farm and parity is significant for lactation persistence, the effect of parity and farm is significant for peak time and the effect of farm, parity and the interaction between season x farm is significant for peak yield. Finally, GWAS analysis was performed to observe which genes were significantly associated with those parameters.

Several genomic regions and candidate genes were identified for lactation persistence, which are widely distributed across autosomal chromosomes, especially on BTA3, BTA9, BTA14, and BTA15.

For peak time, four genomic regions and candidate genes were identified, distributed across autosomal chromosomes particularly on BTA6, BTA7, BTA17 and BTA18.

For what concern peak yield, four regions and candidate genes were found, distributed on autosomal chromosomes specifically on BTA1, BTA4, BTA13 and BTA15.

Numerous candidate genes were found: a total of 111 genes were found for potentially coding for the three analyzed traits highlighting the complex and polygenic nature of them. Lastly, a genome ontology analysis was performed to detect which pathways are commonly shared among the significant genes. Overall, pathways related to Beta1, Beta2 and Beta3 adrenergic receptor signaling pathway seems to play an important role. Interesting to notice is also that the 5HT4 type receptor mediated signaling pathway which mediate serotonin release seems to have a connection with the studied traits. In conclusion, the findings might help to further characterize the molecular mechanisms behind the phenotypic expression of lactation persistence and milk production traits such as peak time and peak yield. This result can be useful to improve the genomic evaluation of those economically relevant traits in the Holstein cattle, starting to introduce these traits into genomic indices for the selection of the best individuals.

Introduction

General

Over the past two decades, dairy cattle farming in Italy has undergone a major change seeing production increase from a total of 9,984,801 tons in 2002 (ISTAT, 2002) to 12,556,459 tons in 2020 (ISTAT, 2020). In contrast with the number of dairy cattle, again at the Italian level, which has decreased from 1,771,899 (ISTAT, 2000) to 1,638,382 (ISTAT, 2020) considering the same period.

This is mainly thanks to improvement in genetic selection and improved nutrition (Sorensen et al., 2008). The increasingly massive use of sexed semen, especially on heifers and later cows as well, has made it possible to speed up genetic progress, ensuring a constant number of female heifers and giving the breeder the opportunity to choose the best animals to replace the culled cows (Holden and Butler, 2018).

The availability of better-quality feed, due to the modernization of agriculture, as well as new discoveries in animal nutrition, have led to the development of rations that are increasingly precise and also consistent in terms of chemical-physical and nutritional characteristics. All this results in faster growth of replacement animals and at the same time in better milk yields of production animals.

Thus, this has caused a significant change regarding what are the potentials and the problems faced by modern livestock farms. Modern animals are more productive, providing higher milk yields, resulting mainly from higher peak lactation and greater persistence.

At the same time, this "productive overbearingness" results in reproductive, metabolic, and immunological imbalances that, if not managed optimally by adapting husbandry practices to the needs of these animals, could reduce the profitability derived from them (Butler, 2000).

For these high-yielding cows, the extended lactation is a great way that allows them to express all the genetic potential, bypassing the problem of "productive overbearingness" and maximizing the profitability.

Transition period

The transition period, generally defined as the period from 3 weeks pre calving to 3 weeks post calving, is a determining factor in the productivity and profitability of a dairy cattle herd. During this period, animals undergo great metabolic and physiological changes in a very narrow time frame.

In fact, it is considered the most delicate period of the entire lactation, as during these 21 days, the risk of diseases that may compromise future milk production is the highest.

In addition, the period from onset to peak production is when the largest number of cows are culled.

What specifically occurs is an alteration of immune and innate host resistance mechanisms of the dairy cow that normally begins approximately 3 weeks prior to calving, is maximal at parturition, and continues until 3 weeks after calving. For this reason, the host defense may be depressed, and disease incidence is high.

This is due to polymorphonuclear neutrophilic leukocyte (PMN) function impaired during peripartum and peripheral blood lymphocytes (PBL) who have a functional role with local protection of the mammary gland (Mallard et al., 1998). This explain why the incidence of environmental mastitis is greatest around parturition (Smith et al., 1985).

Concurrent of alterations in host defense mechanisms is the influence of hormonal balance. Initial investigations of peripartum cows reveal positive relationships between growth hormone kinetics and profiles of anti- body response (Mallard et al., 1998).

Another crucial factor is nutrition. In this phase, animals go through a period when the demand for energy increases markedly, due to parturition, energy used for the development of the mammary apparatus, and later, lactation.

All of this occurs while the ingestion capacity (DMI), and thus the resulting nutrient intake, is reduced and thus unable to meet the animal's nutritional needs (Drackley, 1999).

Requirements for net energy and metabolizable protein by healthy cows at 4 d postpartum exceeded intakes by 26% and 25%, respectively. Furthermore, calculated utilization of net energy and metabolizable protein by the mammary gland for milk production accounted for 97% and 83%, respectively, of intakes, leaving little to supply maintenance needs (Bell, 1995).
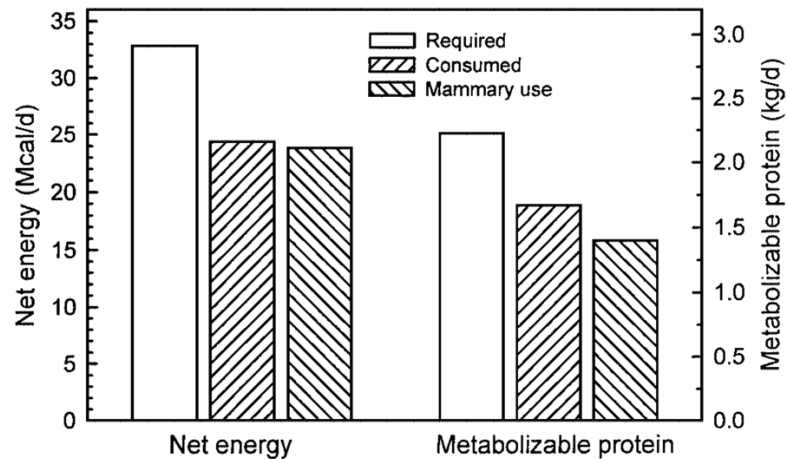
Figure 1: Calculations of amounts of NE and metabolizable protein required, consumed, and utilized by lactating mammary gland of healthy dairy cows at 4 d postpartum (Bell, 1995).(Bell, 1995).

What emerges from all this is how delicate this period is and how much it can affect future lactation if something goes wrong.

During the transition period, milk fever, retained fetal membranes, metritis and displaced abomasum occur with greater incidence, resulting in increased costs due to drugs and veterinary services, as well as increased in-farm labor.

It is easy to see how important the transition period is and how much future lactations depend on its success.

We also have to bear in mind that the risk related to this period is always present, due to for the physiologic changes that animals face at the beginning of lactation.

In this regard, the possibility of extended lactation, could be a feasible alternative in order to decrease the incidence of these periods on the productive life of the animal and ensure better life expectancy.

If this is implemented by choosing high-yielding dairy cows with high persistence, the profitability of the farm will be adequate and, at the same time, the number of transition cows per year will be reduced.

Lactations compared

The duration of a standard lactation is 305 days, that are 10 months.

To date, however, dairy cows are able to maintain milk production for more than 305 days (Österman and Bertilsson, 2003).

Lactation that extends beyond 305 days, referred to as extended lactation, can also be considered advantageous for intensive breeding realities, as long as animals with high lactation persistence are selected (Arbel et al., 2001).

It should be premised that primiparous cows have the best persistence rate, compared with multiparous cows (Arbel et al., 2001).

However, we should also take into account that a primiparous cow produces about 80% of a multiparous cow for the same number of days in milk.
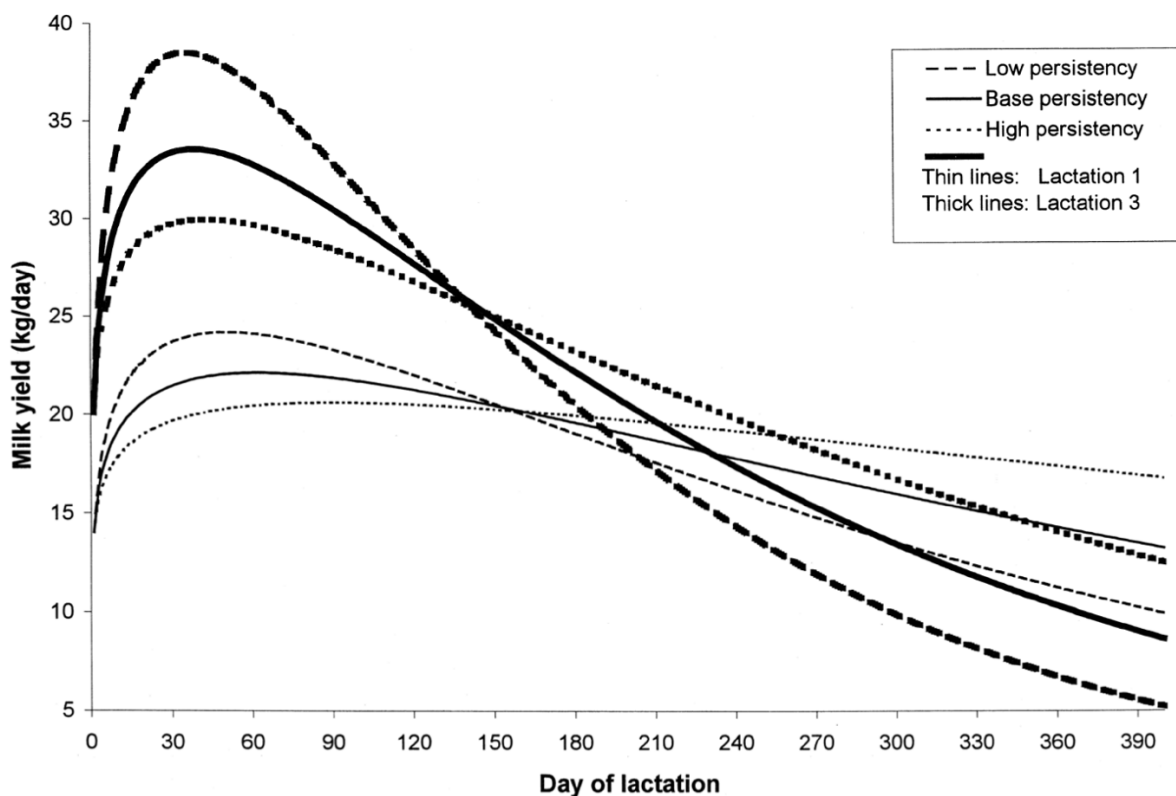

Figure 2: Shape of lactation curve for average producing cow in lactation 1 "thin lines" or lactation 3 "thick lines", for low, average, and high persistence (Dekkers et al., 1998).

Regarding milk components, it is interesting to observe how the trend of milk solid components changes during lactation, continuing beyond the 305 days of standard lactation, in order to observe the trend of these parameters in a hypothetical extended lactation.

Data from the experiment performed by (Sorensen et al., 2008) where multiple groups of cattle went through one or two extended lactations are used as a reference to demonstrate how persistence is moldable and changes depending on how certain conditions related to herd management are handled.

In this experiment, cattle were randomly assigned to two different groups (supplementation with 18% crude protein (CP) concentrates, no supplementation) and each cow was milked two or three times daily, dividing the udder into two halves (diagonally opposite quarters).

Data from cattle managed with concentrate supplementation and three daily milking (managerial management that ensured the best persistence) are taken as reference. Regarding protein, we can observe that CP (%) increased from 3.3% to 3.9% at 420 days, at the same time, the fraction of caseins (measured as a percentage of total protein) remained almost unchanged.
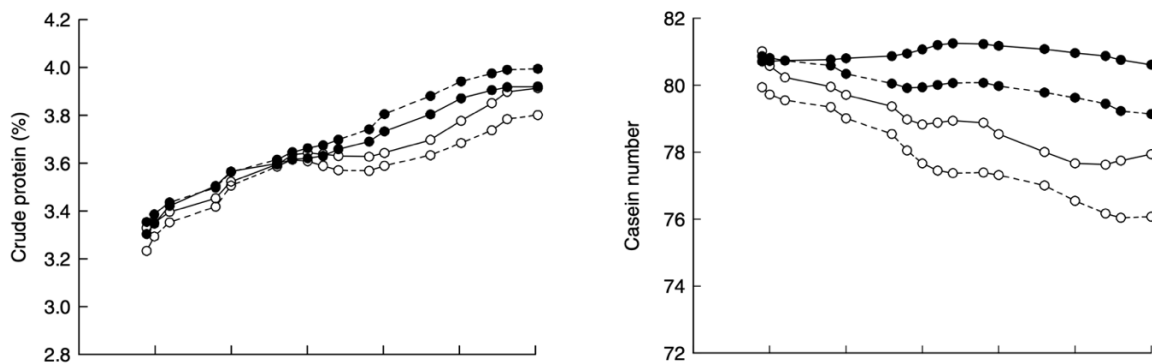

Figure 3: Protein % and casein as a percentage of total protein (full circles and continuous line refers to supplemented diet and three milking/day group) (Sorensen et al., 2008).

As for the percentage of fat, the trend was up and down but still increasing, starting from a Fat (%) of 3.6% to 4.2% at 420 days.

9

Figure 4: Fat % (full circles and continuous line refers to supplemented diet and three milking/day group) (Sorensen et al., 2008).

Lactose had a negative trend, having a decline that was, however, less influential in animals managed in this way. Initially, the values were slightly above 5% to end at 420 with values of 4.6%.

Figure 5: Lactose % (full circles and continuous line refers to supplemented diet and three milking/day group) (Sorensen et al., 2008).

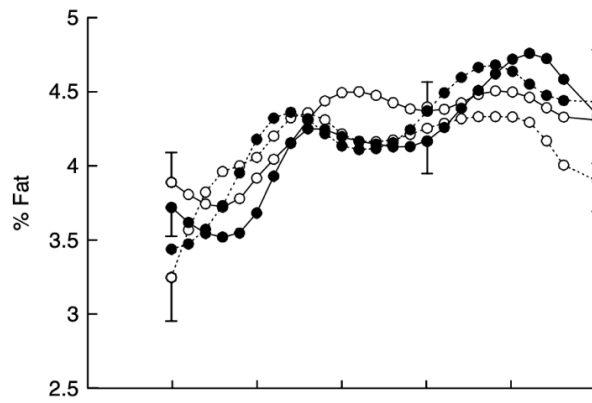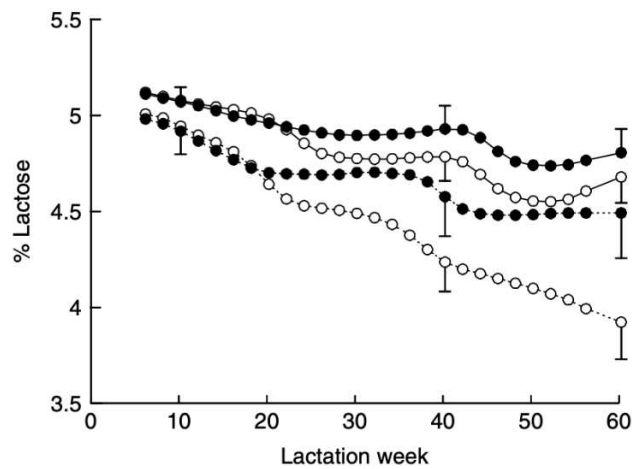Finally, looking at somatic cells, it can be observed that the value of somatic cells always remained below 100,000 CS/ml.
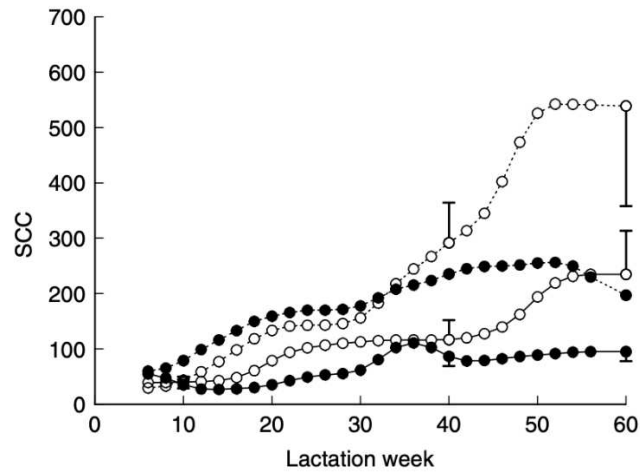
Figure 6: SCC (full circles and continuous line refers to supplemented diet and three milking/day group) (Sorensen et al., 2008).

This finding is also strongly correlated with concentrate supplementation and milking frequency.

The value of caseins and the number of somatic cells demonstrate how the udder managed to remain healthy throughout lactation retarding and limiting the process of apoptosis.

This physiological process leads to a deterioration of milk quality by reduction of casein quantity due to the action of cellular enzymes.

These cell populations, which are naturally present in serum, cross tight junctions that begin to become inefficient and are the main causes of milk quality deterioration as they contribute to raising the SCC and at the same time degrade the casein fraction.

This is an important index of epithelial integrity, which is crucial in the management of extended lactation and, also, a parameter to be taken into account precisely because of the economic outcome.

In addition, an increased frequency of lactation, increases the production of growth hormone (GH) and prolactin, which stimulate local secretion of IGF-1, which has a cell-survival factor action on mammary secretory cells (Sorensen et al., 2008).

Below, it is good to talk about how the dry period should be managed in an appropriate way during extended lactation.

It is often criticized that with extended lactation, animals are transferred to the dry period well before 60 days to calving because milk production falls below profitable levels for the farm.

This occurs when animals are not properly managed for extended lactation, but rather extended lactation is chosen as a strategy because of an extended calving-conception phase due to low fertility of these animals.

In these situations, attention should be paid to the body condition of the animals, as it may compromise the following lactation.

In fact, fat animals are more prone to metabolic disorders in the period following parturition, as well as having lower DMI and lower milk production (Roche et al., 2009).

It is worth mentioning that for every 5 kg of milk produced at dry off, the risk of IMI (intramammary infections) increases by 77 % (Rajala-Schultz et al., 2005).

Non genetic factors affecting persistence

Before analyzing the genetic background of persistence, it should be premised that, the persistence of a dairy cow is affected by a set of factors attributable to the management of the animal itself, which differ from barn to barn and can be defined as the environment.

This set of practices, therefore, allows or not the animal to reach its full potential in terms of persistence.

Increased milking frequency results in increased milk production; it is well known that the third milking adds about 10%-15% milk.

But that's not all, milking frequency also induces greater persistence of lactation itself, which is maintained throughout lactation, even when milking frequency is reduced (Bernier-Dodier et al., 2010).

The reason behind is related to what happens at the hormonal levels, in fact frequent lactation is able to increase the exposure of the mammary epithelium to circulating prolactin, induced by the milking itself, although it is not able to increase basal levels (Hale et al., 2003).

Another key aspect relates to concentrate supplementation, since it was evident that an administration of concentrates, mainly high CP%, improved persistence and udder health (Sorensen et al., 2008).

Another important aspect is related to the voluntary waiting period before breeding the animals, it is attested that a voluntary waiting period of additional 60 days (beyond the conventional 90 for primiparous and 60 for multiparous), results in an increase in milk production, slowing the decline in the latter part of lactation. At the same time, this results in increased production in the following lactation. This effect is more pronounced in primiparous

cows, which by nature have greater persistence, but it is also clearly visible in multiparous cows where it still remains economically beneficial, with a gain of $0.19 and $0.12 per day, respectively, compared with the control group (120 d for primiparous and 70 d for multiparous) (Arbel et al., 2001).

Persistence related genes

Lactation persistence (LP) is defined as the ability of a cow to maintain milk production at a high level after reaching the milk production peak, and this greatly impacts the economic return of the dairy sector. Nowadays more emphasis is being placed on persistence, as this allows increased overall milk production without increasing the risk of metabolic disease during the negative energy balance phase as the peak milk production do.

Indicators of LP based on lactation curves have been studied from a quantitative genetics point of view and it has been found that most of them has a moderate heritability. (Pedrosa et al., 2021) have found that different indicators of LP have been proposed over time, with heritability estimates ranging from 0.14 to 0.24, leaving open possibilities to include those parameters in breeding programs. On top of that, thanks to the advent of the use of genomics in animal genetics, we are nowadays able to study those parameters also from a molecular point of view.

Thus, it is important to define what is a SNP (Single Nucleotide Polymorphism) and what is a GWAS (Genome Wide Association Study) and why it is nowadays highly common, especially in the dairy context, to impute genomic information based on the SNP chip data available.

A SNP is a DNA sequence variation that occurs when a single nucleotide (adenine, thymine, cytosine, or guanine) in the genome sequence is altered, and the alteration is present in at least 1% of the population. A single nucleotide polymorphism (abbreviated SNP) is thus a genomic variant at a single base position in the DNA.

Most of the data from animals that are available and used in the routine genomic evaluation, are SNP chip with low, medium, or high density (in the range of 3K to 800K SNPs), because this is cheaper than performing WGS (Whole Genome Sequencing).

The WGS, also known as full genome sequencing, is the process of determining the entirety, or nearly the entirety, of the DNA sequence of an organism's genome at a single time.

Generating WGS data for many animals is still expensive, but a technique known as genotype imputation can be employed to impute missing markers from animals genotyped with

medium- or high- density genotyping arrays to WGS with accuracies greater than 90–95% (Larmer et al., 2017). Imputation can also be used to allow the use of medium density SNPs information from a pool of animals with only low-density SNP chip which is still the current scenario especially for dairy cows.

The potential of imputed WGS (iWGS) data to discover genetic variants in GWAS has been shown in previous studies of dairy cattle.

Genotype imputation is now an essential tool in the analysis of genome wide association scans. The technique allows geneticists to accurately evaluate the evidence for association at genetic markers that are not directly genotyped. Genotype imputation increases power of genome wide association scans and is particularly useful for combining the association scan results across studies that rely on different genotyping platforms.

The studies taken under observation, started from medium-high density SNPs, from which the genome was imputed with specific software. The studies were carried out by working with IWGS-GWAS. In addition, GO (Gene Ontology) enrichment analyses were carried out to observe the effect of certain genes and the implication of these on lactation persistence (LP).

Regarding the study carried out by (Pedrosa et al., 2021), it was found that BTA18 and BTA28 (where the BTA stands for Bos Taurus Autosome) were the chromosomes with the most significant regions. The genes highly associated to LP on BTA18 were *ARHGAP35*, *NPAS1*, *TMEM160*, *ZC3H4*, and *SAE1* and on BTA28 were *ZMIZ1* and *PPIF*. None of these genes were previously linked to LP.

Another important finding was the gene network connection among *ARHGAP35*, *TMEM160*, and *SAE1* (BTA18) and *ARHGAP39*, *PPP1R16A*, *FOXH1*, and *CYHR1* (BTA14). No other study has reported a close network interaction between some of the main genes responsible for LP and milk production traits, indicating potential pleiotropic effects.

Three GO terms were significantly enriched for LP, (GO:0030334) (regulation of cell migration), GO:19003955 (positive regulation of protein targeting to mitochondrion), and (GO:0004867) (serine-type endopeptidase inhibitor activity). The genes linked to the GO terms *ABI3* and *LDB2* (GO:0030334), *SAE1* (GO:1903955), and *SERPINB6* and *SERPINB9* (GO:0004867) could also be considered as novel candidate genes and its molecular role related to LP should be deeper investigated.

Fifteen genes were significantly associated with LP and at least one of the milk production traits, as *CXCL13* and *LDB2* (MILK); *ZMIZ1*, *bta-mir-371*, *NLRP12* and *PPIF* (FAT); *INPP5A*

(FAT%); *SERPINB6*, *SERPINB9*, *IGF2BP1* and *DLX4* (PROT%). Additionally, LP, FAT%, and PROT% had common candidate genes between the three traits simultaneously: *ABI3*, *GNGT2*, *B4GALNT2*, and *PHOSPHO1,* demonstrating their importance on both milk production, but also in the duration of peak lactation.

Another recent study made by (Do et al., 2017), highlighted several BTA significantly associated with LP highlighting the complexity and polygenicity of the parameters associated to LP.

An interesting region detected in this study was around 107 Mb on BTA5, where the most significantly associated SNP (ARS-BFGL-NGS-2399) and several suggestive SNP were located. The SNP ARS-BFGL-NGS-2399 is in an intronic region of tetraspanin 9 (*TSPAN9*), a gene member of the tetraspanin family known to play important roles in cell development, activation, growth, and motility by mediating signal transduction in mammals.

On BTA2, the *MAN1C1* gene contains a significantly associated SNP for LP in its intronic region. On BTA9, a significant SNP (Hapmap54471- rs29017484) was in the intronic region of *MAP3K5* gene.

More than 100 QTLs for milk yield and milk component traits have been reported around the region of BTA14 which was also significant in this study, and where the acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) gene is located.

However, a significant SNP (ARS-BFGL-NGS-70879, position: 50,596,757) for LP identified in this study is at ~50 Mb from the *DGAT1* gene but close to a previously reported QTL (at 48.9 Mb) for LP on BTA 14.

A significant SNP (ARS-BFGL-NGS-107174) on BTA 19 detected for LP is located in the intronic region of *TEX14*, which is known to play a role in intercellular bridges in germ cells.

On BTA 20, 3 significant SNP (ARS-BFGL- BAC-27930, Hapmap48202-BTA-118947, ARS-BFGL-NGS-114946) for LP are in the intronic regions of the *HCN1*, *MRPS30*, and *CCL28* genes, respectively.

In addition, growth hormone receptor (*GHR*) and 3-hydroxy- 3-methylglutaryl-coenzyme A synthase 1 (*HMGCS1*) genes are located close to the significant SNP on BTA 20.

For what concern the enriched pathway, the most significant biological process (GO:0045723) enriched for the LP gene list was "positive regulation of fatty acid biosynthetic process".

(GO:0006629) term "lipid metabolic process" is enriched with *CEL*, *PAFAH2*, *CPNE6*, *HMGCS1*, *PAFAH1B2*, *GDPD1*, and *SREBF2* genes, and some of them are known to be involved in milk production and the lactation process.

Another notable biological process for LP was (GO:0007595) term "lactation" enriched by 4 genes (*GC*, *HK2*, *CSN3*, and *CSN2*) from the input gene lists. Both β-CN (*CSN2*) and κ-CN (*CSN3*) genes are vital for milk production because they directly regulate milk protein synthesis.

Two GO terms "positive regulation of intrinsic apoptotic signaling pathway" (GO:200124) and "intrinsic apoptotic signaling pathway in response to DNA damage" (GO:0008630) related to the mechanism of apoptosis were identified for LP.

It is also worthwhile mentioning a significant biological process, "insulin receptor signaling pathway" (GO:0008286), which defines the series of molecular signals generated as a consequence of insulin receptor binding to insulin.

The GO term, "protein ubiquitination involved in ubiquitin-dependent protein catabolic process" (GO:004278) was also enriched for LP.

To sum up, we can conclude that from the literature review emerged the potential to improve LP from a genetic point of view since the moderate heritability of parameters associated to LP. However, the GWAS studies highlighted the complex nature of those traits with several genes potentially associated to those parameters.

Material and methods

Description of farms

Two farms were used for this study and are described below.

Refer to appendix at bottom for more information on farms

Az. Agr. Volpere

The farm "Az. Agr. Volpere" is located in Remedello, in the province of Brescia.

The farm consists of a single farm body. The number of lactating animals is currently about 150 cows. The 97% of the cows are Italian Holstein, who is the only breed used for this experiment.

The farm produces drinking milk, but it's also qualified for the production of Grana Padano cheese.

The lactating cow barn is newly built in 2019 and has various amenities such as showers, nebulizers, destratifiers and fans.

There are also older barns, where heifers, dry cows and calves are kept.

The milking parlor is herringbone, 10 + 10 places, and two milking are done daily.

The replacements of culled cows are in-house, although high genetic value animals are sporadically purchased.

All animals on the farm have been genotyped although with different SNP panels (ranging from low density – 3K to medium density 50K).

Az. Agr. Canili

The farm "Az. Agr. Canili" is located in Bozzolo, province of Mantua.

The farm consists of a single farm body similarly to Az. Agr. Volpere.

The number of lactating animals is currently about 720 cows in lactation, therefore a farm with over 4x more lactating animals compared to the previous described farm. In this farm the 99% of the cows are Italian Holstein, and only them are used for this experiment.

The farm produces high quality drinking milk.

Within the farm there are several barns, some older and some modern.

The last major expansion took place in 2020 where the new rotary milking facility and new barns for lactating cows were built. New barns have a capacity of 500 cows.

Some of the lactating cows are still maintained at the previously built concrete and brick barns.

The rotary has 60 milking places and three milking per day are done.

Replacement for culled cows is in-house.

All animals in the herd have their genomes registered.

During the collection of the samples (biennium 11/2019-11/2021), the movement of animals to the new barns was happened, as well as a major increase in the number of animals.

On April 29, 2021, the rotary came into function, and for the first week, the animals were milked both in the rotary and in the previous milking facility, which is a Delaval facility at 75 degrees 14 + 14 places. Also, in this farm the majority of the animals have been genotyped using several SNP panels.

Test day milk check

The test day controls used for this study are those collected by the Lombardy Regional Breeders Association (ARAL). For each farm, 18 functional controls were collected over two years. For the "Az. Agr. Volpere" the dates of checks are shown in the table below (Table 1).

Table 1: General overview of the data collected in terms of time frame for the Az. Agricola Volpere

Az. Agricola Volpere (Cioli)									
Functional controls									
2019	Date of the functional control:	29/11/19								
	Days between this control and the following:	53								
2020	Date of the functional control:	21/01/20	24/02/20	28/04/20	08/06/20	14/07/20	24/08/20	30/09/20	05/11/20	11/12/20
	Days between this control and the following:	34	64	41	36	41	37	36	36	38
2021	Date of the functional control:	18/01/21	23/02/21	31/03/21	06/05/21	08/06/21	15/07/21	31/08/21	05/10/21	10/11/21
	Days between this control and the following:	36	36	36	33	37	47	35	36	

For "Az. Agr. Canili" the dates of checks are shown in the table below (Table 2).

Table 2: General overview of the data collected in terms of time frame for the Az. Agricola Canili.

Az. Agricola Canili (Chizzoni)									
Functional controls									
2019	Date of the functional control:	26/11/19								
	Days between this control and the following:	49								
2020	Date of the functional control:	14/01/20	17/02/20	28/04/20	03/06/20	06/07/20	06/08/20	16/09/20	19/10/20	25/11/20
	Days between this control and the following:	34	71	36	33	31	41	33	37	57
2021	Date of the functional control:	21/01/21	22/02/21	26/03/21	27/04/21	03/06/21	07/07/21	15/09/21	22/10/21	25/11/21
	Days between this control and the following:	32	32	32	37	34	70	37	34	

Phenotype Data filtering

Two datasets containing milk test day data were available from the two different farms. The test day data were recorded between 2019 – 2021 with a total of 12,993 records belonging to 1035 cows. The following filters were applied in R software (R Development Core Team, 2011):

Lactations were divided in four classes based on parity.

Test day data were kept if recorded from 5 to 305 days since it is known that milk controls recorded after 305 days are more difficult to model which might compromise the whole lactation evaluation (Bakri et al., 2022).

Lactations were required to have at least five test day records for each animal within lactation to ensure an accurate estimation of the lactation curve parameters and to exclude animals with extremely short lactations.

Outliers were defined as value above or below the mean ± 3SD and were removed.

Fitting the lactation curve

To describe the lactation curve and determine production characteristics as Peak Yield (PY), Peak time (PT) and Persistence (Per), the Wood's incomplete Gamma function (WOOD, 1968) was fit to the test day yields:

$$Y_t = a * t^b * \exp^{-ct}$$

Where Y_t = The test day milk, protein, fat yield at time t; t = The day of lactation; a = The parameter associated with the beginning of lactation; b = The parameter associated with the increasing phase before peak yield; c =The parameter associated with the decreasing phase during the declining phase until the end of lactation; exp: The base of the natural logarithm.

Based on the curve parameterization, the following production characteristics were calculated (Tekerli et al., 2000):

Peak yield (PY) as $PY = a * \left(\dfrac{b}{-c}\right)^b * exp^{-b}$

Peak time (PT) as: $PT = \dfrac{b}{-c}$

Persistence (PER) as: $Per = -(b + 1) * Ln(-c)$

ANOVA

A mixed model in Asreml-SA 4.1 was used to assess the effect of animal and environmental factors on the three above mentioned production characteristics (Peak yield, peak time and persistence) for the trait: kg of milk.

The factors included in the model were:

Parity class: first, second, third and over four parities

Season of calving (four classes)

Farm (two classes)

The interaction between the above-mentioned factors was also considered in the model.

Factors with a p-value lower than 0.05 were considered as significant.

Genome Wide Association studies

For all cows included in this study, genotype data in the form of single nucleotide polymorphism (SNPs) were available. Those cows were genotyped with several different panels ranging from low density (3K) to medium density (50K). The animals genotyped with low-density panels were imputed to medium density (85K) using PedImpute as routinely performed by ANAFIBJ (National Association of breeders of Holstein, Jersey and Brown Swiss cows). To guarantee high accuracy during the imputation pipeline, females were retained for this study only when both sire and sire of the dam were (i) genotyped and (ii) used in the imputation pipeline. Quality control (QC) excluding poorly genotyped and faulty data was performed on the 29 autosomal chromosomes by using PLINK v1.90. The QC was based on the following criteria: call rate of <90%, parent–offspring SNP mismatch of <0.01, minor allele (<0.01) and genotype (<0.001) frequencies, and extreme deviation from the Hardy–Weinberg equilibrium (P < 0.0005).

The animals and SNPs that passed the QC were further used in the GWAS performed in the StatGenGWAS package in R software. In the GWAS we corrected for the above-mentioned factors included also in the ANOVA as well as for the genomic relationship matrix. Two significant thresholds were used in the GWAS a Bonferroni threshold (0.05/number of SNPs passing the QC) and a suggestive threshold of -log10(p-value) equals to 4.

The ARS-UCD1.2 - genomic coordinates of the region containing significant SNPs (± 50 Mbp) were used to retrieve candidate gene lists and annotations from the Biomart web interface in Ensembl release 94 (Aken et al., 2016). In addition, the potential overlaps with QTLs, present

in the Cattle QTL database (Hu et al., 2022), were considered. Finally, statistical overrepresentation test of pathways (GO terms) of the candidate genes was conducted using PANTHER 17.0. The level of significance for the overrepresented pathways was set based on FDR correction using a Fisher test.

Results

Descriptive statistics and QC of the genotype data

Number of test-days (TD), cows and average production and milk quality traits are reported in Table 3. Overall, we found the lowest mean milk yield for primiparous cows and the highest in the tertiparous one. Similar trends were also found when evaluating milk quality traits (mean % fat, mean % protein). The lactation length ranged from 272 days on average in the primiparous to 277.1 in the tertiparous.

Table 3: Descriptive statistics of the datasets used in this study divided by parity class

Variable	Parity 1	Parity 2	Parity 3	Parity ≥ 4
Test-days (TD)	3616	2459	1553	1711
Cows	523	343	216	182
Average n. of TD per cow	6.91	7.17	7.19	9.40
Mean milk yield (kg ± SD)	37.8 (8.25)	44.6 (11.42)	46.2 (12.27)	44.27 (12.95)
Mean % fat (SD)	3.78 (0.75)	3.81 (0.85)	3.87 (0.89)	3.83 (0.95)
Mean % protein (SD)	3.34 (0.31)	3.36 (0.35)	3.35 (0.34)	3.30 (0.36)
Average lactation length	272.0 (29.99)	275.4 (26.70)	277.1 (26.83)	276.13 (29.37)

Quality control of the genotype data

A total of 78,022 SNPs and 775 cows were kept after QC. The average genotyping call rate was 0.99, a total of 107 SNPs were removed due to low coverage in the dataset, 10 SNPs were removed due to HWE and 5540 for the minor allele threshold (MAF).

Wood Parameters

Kg of milk

From the application of the Wood formula three parameters were calculated (peak yield, peak time, and persistence) which are shown in Table 4 for the kg of milk. The highest persistence was found in the primiparous (7.48) and the lowest in the tertiparous (6.95). Regarding secondiparous cows, it is interesting to note that the persistence in terms of milk peak it is 31 days earlier compared to primiparous cows. In third lactations cows we find the absolute

lowest data in terms of persistence, and days in milking at peak lactation (65 DIM), but they are the group with the highest values in terms of peak yield with an average value of 55.22 kg. The graphical representation of the curves is shown in figure 7.

Table 4: Wood Parameters estimation and production related traits divided by parity class for kg of milk

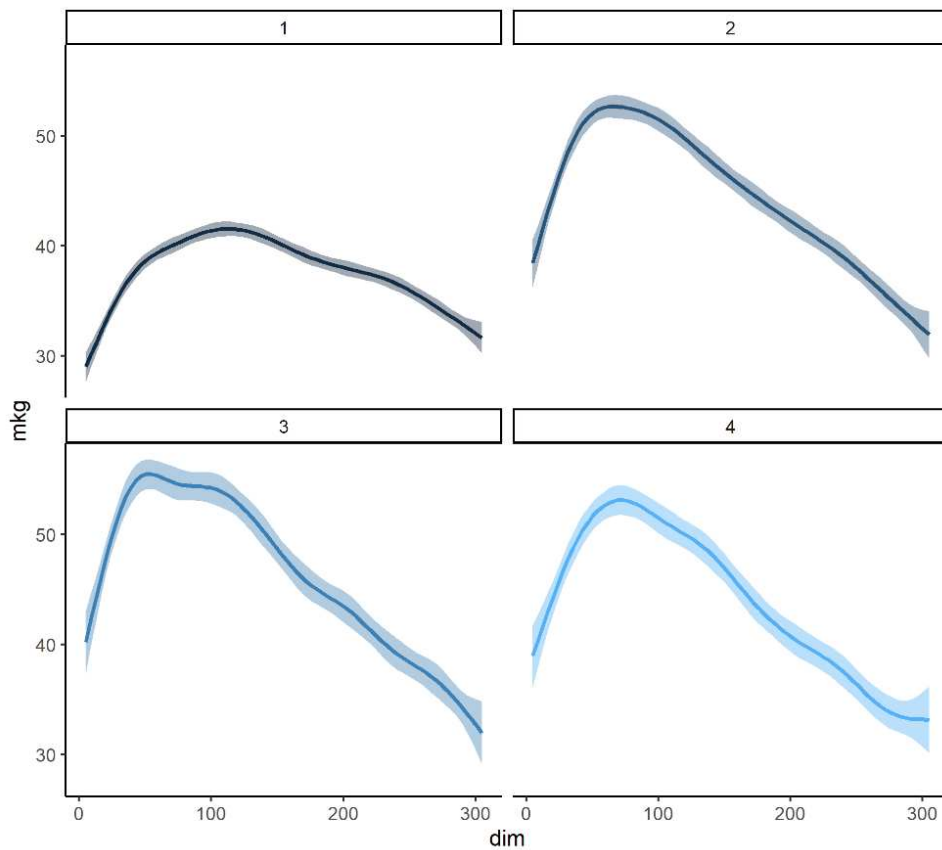| Parameter	Parity 1	Parity 2	Parity 3	Parity ≥ 4
Average a	17.09	23.16	25.19	22.04
Average b	0.24	0.25	0.25	0.27
Average c	-0.0024	-0.0036	-0.0038	-0.0039
Peak Yield (kg)	41.07	52.22	55.22	52.33
Peak Time (days)	100	69	65	68
Persistence	7.48	7.03	6.95	7.03

Figure 7: Modelled lactation curve per kg of milk per day following Wood function divided by parity class.

Kg of fat and Kg of protein

Third lactations cows are by far the category with the highest fat (kg) and protein (kg) production in terms of peak yield (Table 5 and Table 6). Kg of fat and kg of protein show a decreasing trend continuing along the lactation, this is mainly to be attributed to the decline in milk production in kg, being that these data are expressed in kg of fat and protein rather than in % of total milk.

Similarly, to what reported for kg of milk, for fat and protein, the trend is the same: persistence is better in primiparous than in the other groups of animals, however, the peak of fat and protein occurs at 97 DIM and 129 DIM respectively deviating from the milk peak (on average at 100 DIM).

The graphical representations of the curves are shown in Figure 8 and Figure 9.

Table 5: Wood Parameters estimation and production related traits divided by parity class for kg of fat.

Wood Parameters (fat)	Parity 1	Parity 2	Parity 3	Parity ≥ 4
Average a	1.19	1.56	2.02	1.65
Average b	0.05	0.07	0.02	0.06
Average c	-0.0006	-0.0019	-0.0014	-0.0020
Peak Yield (kg)	1.44	1.90	2.07	1.94
Peak Time (days)	97	39	11	32
Persistence	7.90	6.75	6.66	6.63
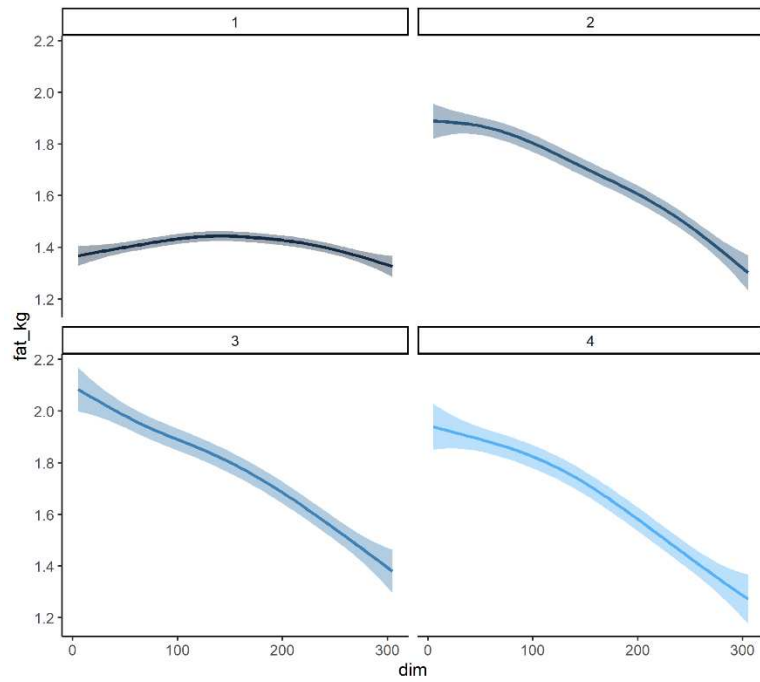
Figure 8: Modelled lactation curve per kg of fat per day following Wood function divided by parity class.

Table 6: Wood Parameters estimation and production related traits divided by parity class for kg of protein.

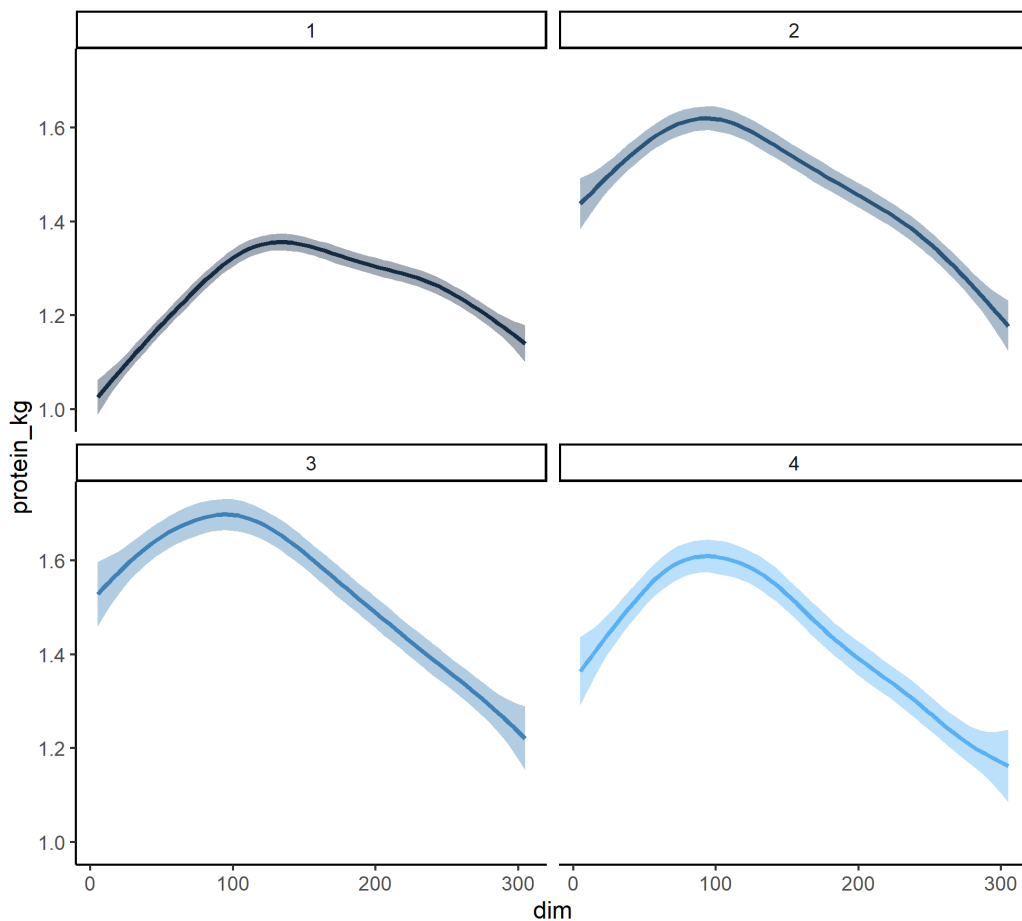| Wood Parameters (protein)	Parity 1	Parity 2	Parity 3	Parity ≥ 4
Average a	0.62	0.99	1.06	0.87
Average b	0.19	0.15	0.15	0.19
Average c	-0.0015	-0.0021	-0.0023	-0.0025
Peak Yield (kg)	1.32	1.62	1.71	1.60
Peak Time (days)	129	71	66	73
Persistence	7.76	7.09	7.00	7.08

Figure 9: Modelled lactation curve per kg of protein per day following Wood function divided by parity class.

Overall, it is possible to observe in the figures above (figure 7, figure 8, figure 9) that primiparous cows are the ones manifesting the best persistence value and at the same time they are the group of animals in which the lactation peak is shifted more than other to the right in terms of days (peak time: 100 DIM). At the same time, it is worth noting that they are the group of animals with the lowest production peak (41.07 kg) as reported in table 4.

As for fat and protein, peak yield is higher in multiparous than in primiparous (it should be remembered that fat and protein are expressed in kg and not in percentage and thus affected by milk yield) Interesting to notice is that the peak of fat production (kg) is in general shifted towards the beginning of lactation especially for multiparous cows. In contrast the peak of kg of protein, is more in line with the peak of milk production (kg).

The last category consists of animals with at least 4 or more lactations and is the category with the most diverse number of individuals. In this category we can observe a slight improvement in persistence (milk) and also days to peak lactation (68 DIM), although production at peak is

slightly lower (52.33 kg) compared to the other class of multiparous cows (parity 2 and parity 3). Regarding fat and protein peaks, in parity 4 the production in kg is lower than in third lactations cows, although it is at higher levels than in primiparous and secondiparous.

ANOVA

Below are presented the results of the ANOVA for the parameters:

Persistence

Peak Time

Peak yield

For each parameter the following results are shown F statistics table, Least square means (LSM) for the significant factors and a graphical representation of the results.

Persistence

In table 7 the results of the ANOVA analyses for the parameter persistence are reported. The factors that were found to be significant for persistence are: season, farm, and parity. The interaction between the above-mentioned factors resulted non-significant.

Table 7: Result of ANOVA analysis of persistence

Results from ANOVA for persistence					
Source of variation	NumDF	DenDF con	F-inc	F-con	P-con
season	3	823.0	3.47	3.59	0.013
farm	1	823.0	12.37	15.33	<0.001
parity	3	823.0	17.97	17.97	<0.001
season.farm	3	823.0	1.8	1.5	0.214
farm.parity	3	823.0	0.2	0.21	0.889
season.parity	9	823.0	0.9	0.9	0.528

Num DF: numerator of degree of freedom; **DenDF con**: denominator of degree of freedom; **F-inc/F-con**: wald F statistics; **P-con**: P-value.

The persistence of the lactation curve is also affected by the calving season; in fact, it is possible to observe that animals that begin lactation in the fall are those that maintain a higher persistence rate throughout lactation (Figure 10 and Table 8).

Table 8 Least square means (LSM) calculated for the parameter persistence, in detail for the effect of season

Season		
season	Predicted Value	Standard Error
summer	7.32	0.09
autumn	7.52	0.08
winter	7.18	0.08
spring	7.16	0.10
SED: Standard error of difference: 0.12		
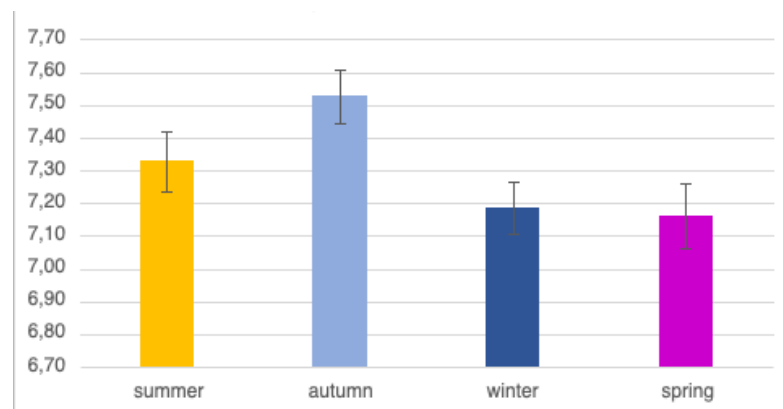
Figure 10: Persistence predicted value per season

Moreover, it is highly visible the difference between the two farms as shown in Table 9 and Figure 11. This could be traced back to the different management of them. In fact, the Az. Agr. Canili farm performs three daily milkings while the Az. Agr. Volpere farm performs only two. This results in a difference in production of about 15% of milk quantity, but not only that, frequent milking, especially in the early part of lactation results in greater stimulation of the mammary apparatus and this results in greater production and persistence as in line with the literature (Bernier-Dodier et al., 2010).

Table 9: Result of ANOVA analysis of persistence, in detail for the effect of farm

Farm		
farm	Predicted Value	Standard Error
Az. Agr. Canili	7.46	0.03
Az. Agr Volpere	7.13	0.08
SED: Standard error of difference: 0.094		
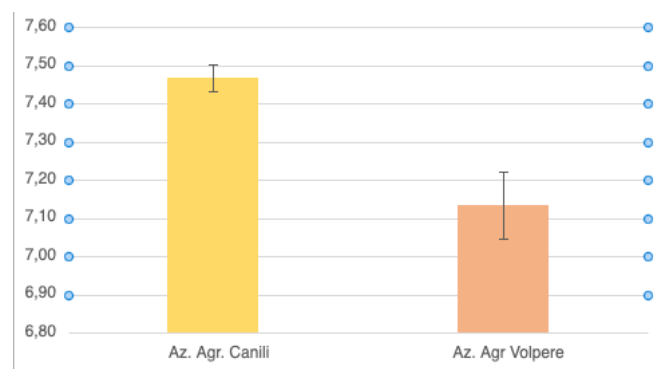
Figure 11: Persistence predicted value per farm

In addition, it is possible to observe how the persistence of the lactation curve is influenced by parity (Table 10 and Figure 12). It is indeed visible that primiparous cows are the animals that absolutely have the highest persistence value. This finding is confirmed by various studies including that of (Arbel et al., 2001).

Table 10: Result of ANOVA analysis of persistence, in detail for the effect of parity.

Parity		
Parity	Predicted Value	Standard Error
1	7.64	0.05
2	7.11	0.07
3	7.17	0.11
≥4	7.27	0.12
SED: Standard error of difference: 0.13		
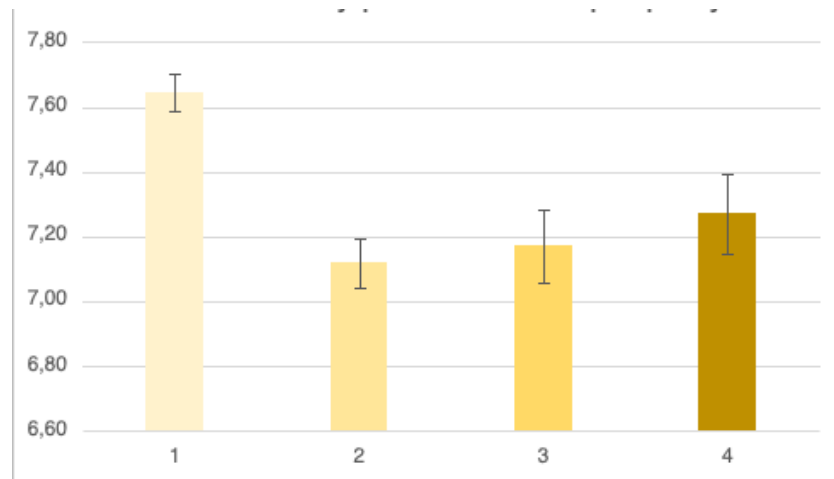
Figure 12: Persistence predicted value per parity

In figure 13 it is possible to visualize the non-significant interaction between parity and season.

Figure 13: Milk production (kg) divided by season and parity

Peak time

The factors that were found to be significant for peak time are: farm and parity. The interaction between the above-mentioned factors resulted non-significant as well as the season (P-value=0.084).

Table 11: Result of ANOVA analysis of peak time

Results from ANOVA analysis of peak time					
Source of variation	NumDF	DenDF con	F-inc	F-con	P-con
farm	1	823.0	5.56	11.13	<0.001
parity	3	823.0	50.49	50.49	<0.001
season.farm	3	823.0	3.95	2.22	0.084
season	3	823.0	2.29	2.12	0.096
farm.parity	3	823.0	1.16	1.15	0.329
season.parity	9	823.0	1.32	1.32	0.221

Num DF: numerator of degree of freedom; **DenDF con**: denominator of degree of freedom; **F-inc/F-con**: wald F statistics; **P-con**: P-value.

It can be observed that there is a difference of about 15 DIM between the two different farms (Table 12 and Figure 14). This finding could be related to a higher frequency of milking, or to a higher number of primiparous cows in one of the two farms.

Table 12: Result of ANOVA analysis of peak time, in detail for the effect of farm

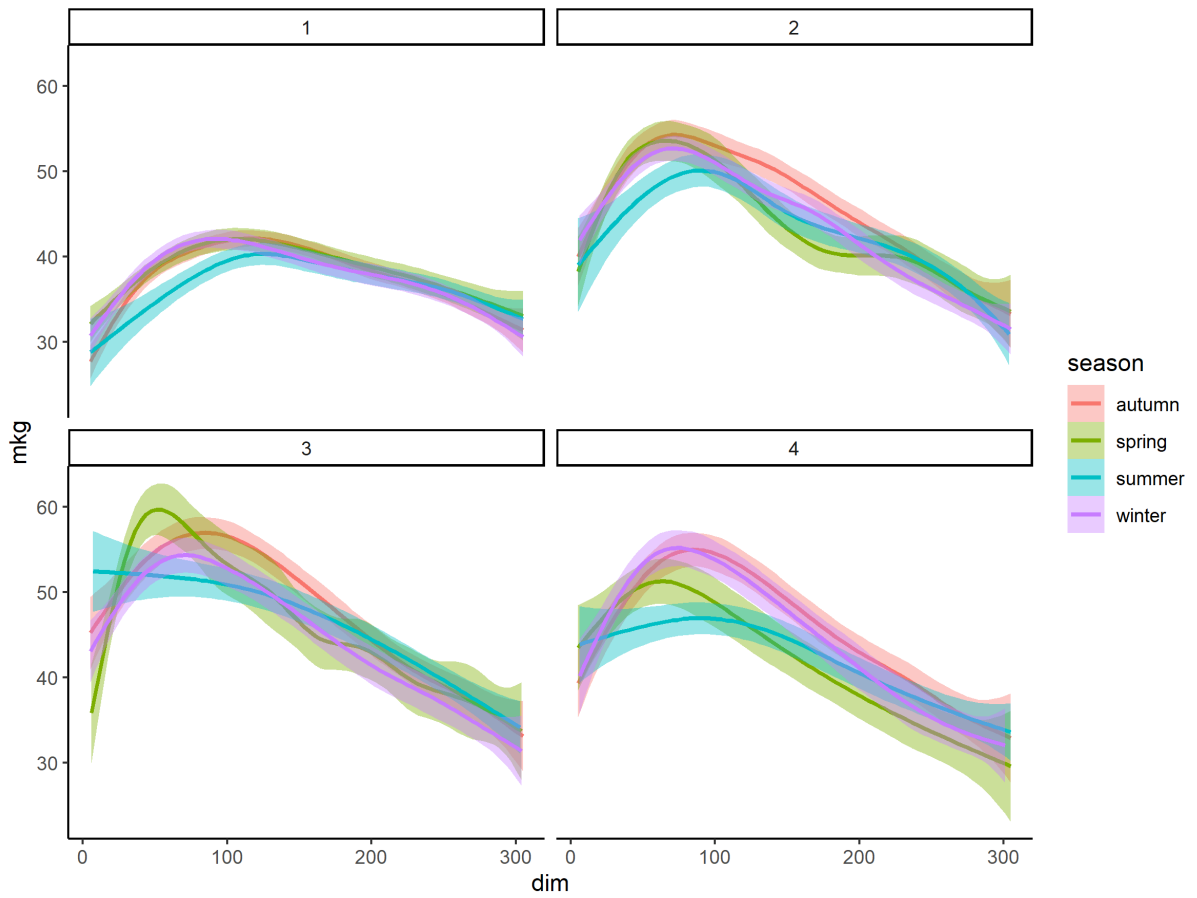| Farm		
farm	Predicted Value	Standard Error
Az. Agr. Canili	80.41	1.55
Az. Agr. Volpere	65.65	3.83
SED: Standard error of difference: 4.12		
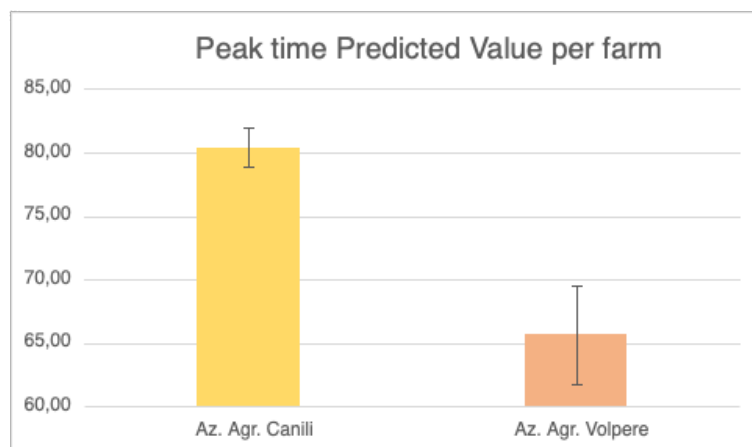
Figure 14: Peak time (DIM) predicted value per farm

Peak time is influenced in an important way by parity (Table 13 and Figure 15). It is observed that days to peak lactation is about 100 DIM for primiparous cows while the value is between 60 DIM and 70 DIM for the other groups of animals. This high value for primiparous cows could partly be related to the higher persistence of these individuals.

Table 13: Result of ANOVA analysis of peak time, in detail for the effect of parity

Parity		
Parity	Predicted Value	Standard Error
1	99.95	2.53
2	63.07	3.33
3	60.85	4.87
4	68.25	5.41
SED: Standard error of difference: 5.97		
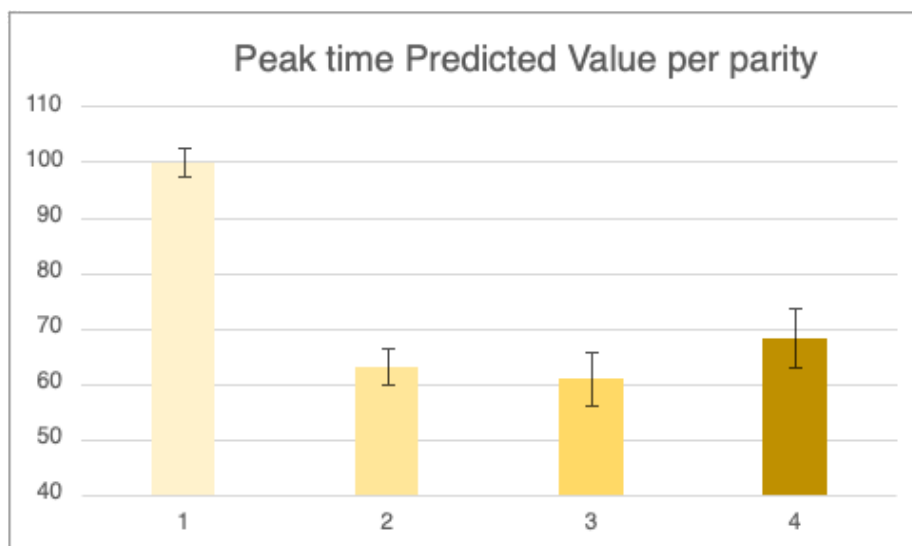
Figure 15: Peak time (DIM) predicted value per parity

Peak yield

The factors that were found to be significant for peak time are: farm, and parity (Table 14). The interaction between the above-mentioned factors resulted significant for season and farm.

Table 14: Result of ANOVA analysis of peak yield

Results from ANOVA analysis of peak yield						
Source of variation	NumDF	DenDF con	F-inc	F-con	M	P-con
farm	1	823.0	171.14	117.80	A	<.001
parity	3	823.0	139.45	139.45	A	<.001
season.farm	3	823.0	7.83	4.13	B	0.006
season	3	823.0	2.21	0.74	A	0.529
farm.parity	3	823.0	0.38	0.47	B	0.704
season.parity	9	823.0	1.45	1.45	B	0.163

Num DF: numerator of degree of freedom; **DenDF con**: denominator of degree of freedom; **F-inc/F-con**: wald F statistics; **P-con**: P-value.

34

It is interesting to observe the significant difference between the two farms, that is evident from the projections of the lactation curves divided by herd; the herd that performs three daily milkings has a higher peak production and also it is maintained for a longer period of time before declining (Figure 16).
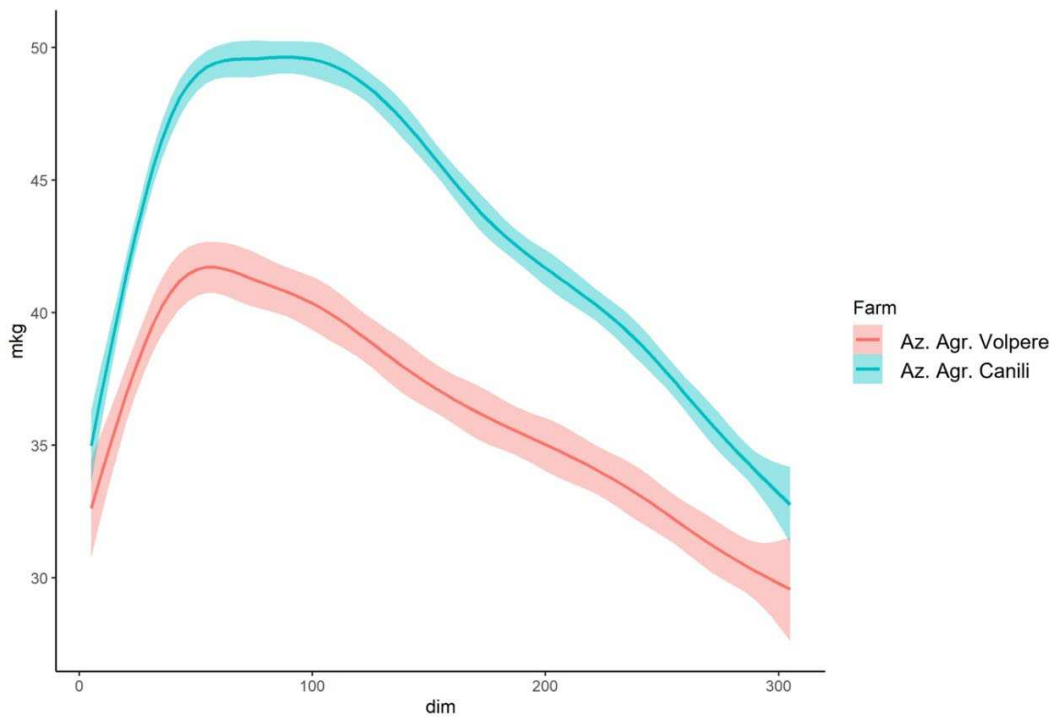
Figure 16: Milk production (kg) during lactation, divided by farm

From the comparison of the curves in the two farms in fat kg during lactation, graphically it seems in contrast that fat benefits less from the three daily milking, in fact, the values are similar for the two farms, with a similar trend during the entire lactation, regardless of the parity order (Figure 17).
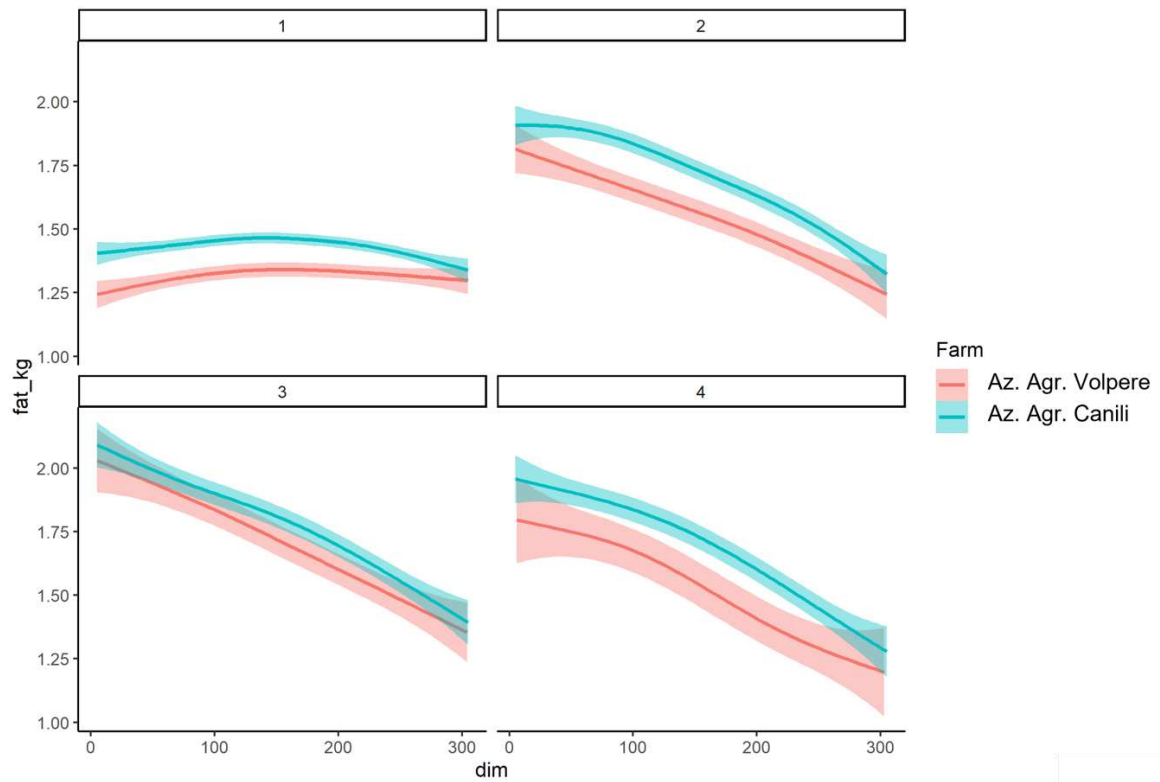
Figure 17 Fat production (kg) during lactation, divided by farm and by parity

Protein trends, on the other hand, differ considerably from the comparison of the two farms, probably rewarding milking frequency. In fact, in the farm where cows are milked more frequently, the production in kg of protein is higher and the trend of the protein curve also differs considerably. This trend occurs indiscriminately across all animal groups (Figure 18).
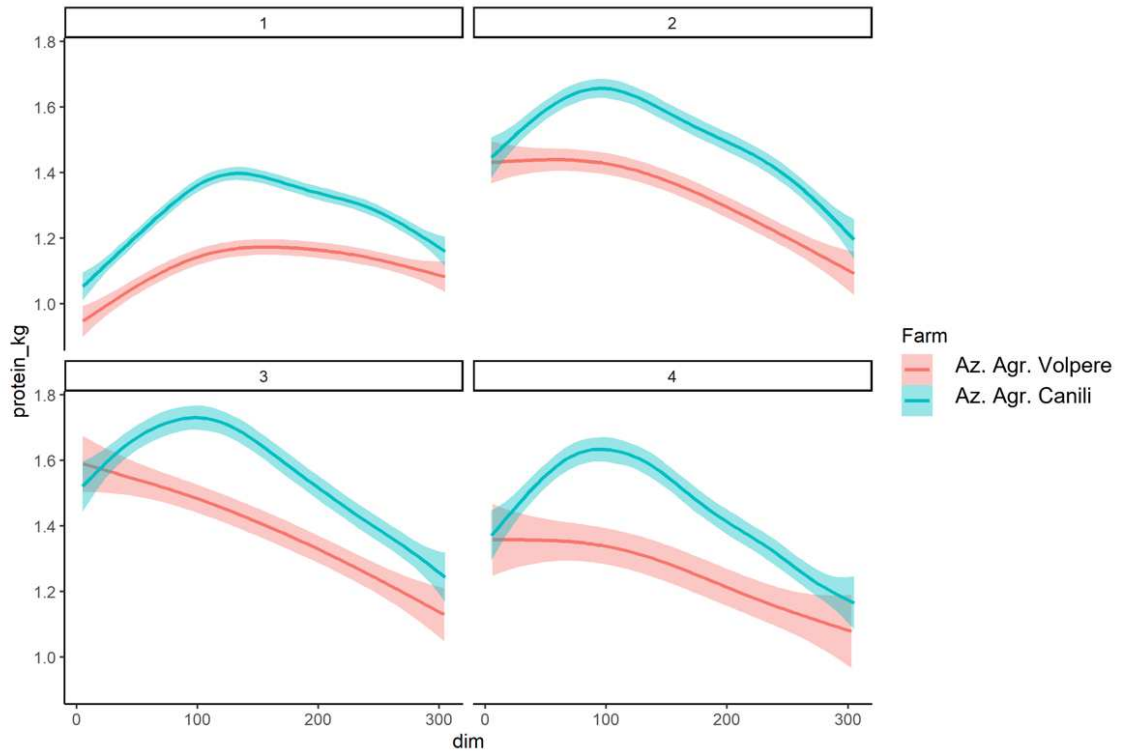
Figure 18: Protein production (kg) divided by farm and by parity

Table 15: Result of ANOVA analysis of peak yield, in detail for the effect of farm

Farm		
farm	Predicted Value	Standard Error
Az. Agr. Canili	53.20	0.32
Az. Agr. Volpere	46.10	0.79
SED: Standard error of difference: 0.85		
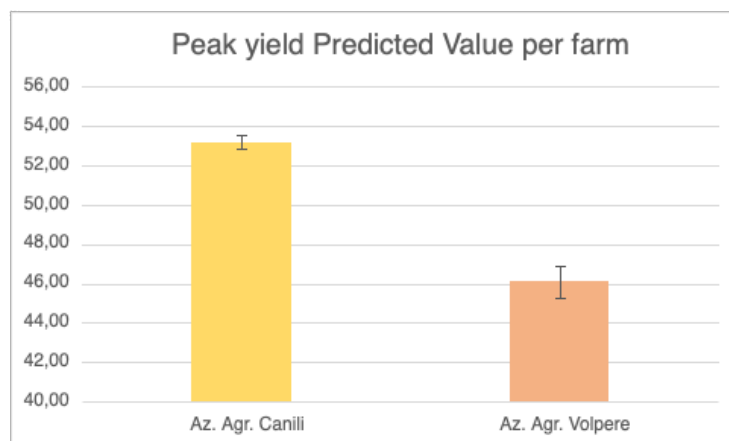
Figure 19: Peak yield predicted value per farm

Peak lactation reaches its highest values in the tertiparous cows group (Table 16 and Figure 20). Primiparous cows have the lowest value of kg of milk at peak lactation, this finding is understandable when considering that a primiparous cow does not have the production capacity of a tertiparous cow and produces about 80% of a tertiparous cow. It is also recognized that there is a negative correlation between lactation persistence and peak lactation (Sorensen et al., 2008).

Table 16: Result of ANOVA analysis of peak yield, in detail for the effect of parity

Parity		
Parity	Predicted Value	Standard Error
1	41.16	0.52
2	49.74	0.69
3	55.45	1.01
4	52.24	1.12
SED: Standard error of difference: 1.24		
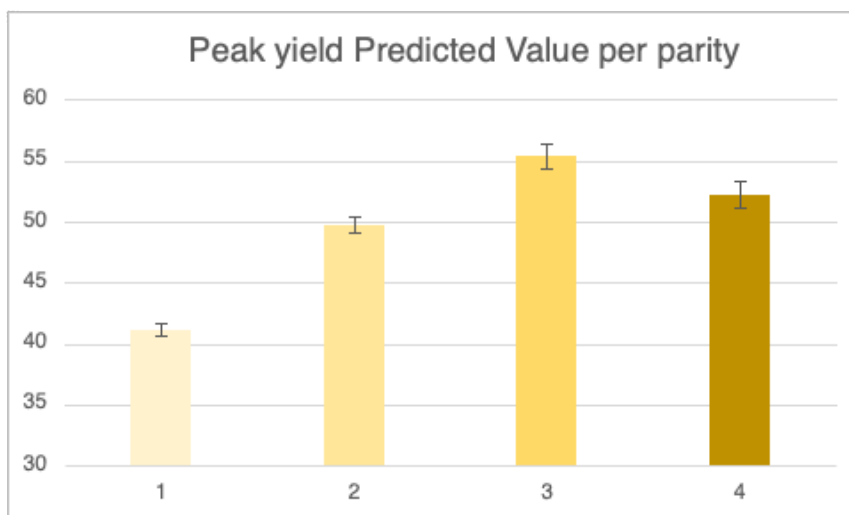
Figure 20: Peak yield predicted value per parity

Table 17: Result of ANOVA analysis of peak yield, in detail for the effect of season.farm

Season x Farm			
Season	Farm	Predicted Value	Standard Error
summer	Az. Agr. Canili	53.03	0.69
summer	Az. Agr. Volpere	43.11	1.46
winter	Az. Agr. Canili	53.76	0.58
winter	Az. Agr. Volpere	48.81	1.31
autumn	Az. Agr. Canili	53.58	0.51
autumn	Az. Agr. Volpere	44.17	1.36
spring	Az. Agr. Canili	52.42	0.74
spring	Az. Agr. Volpere	48.29	1.61
SED: Overall standard error of difference: 1.55			
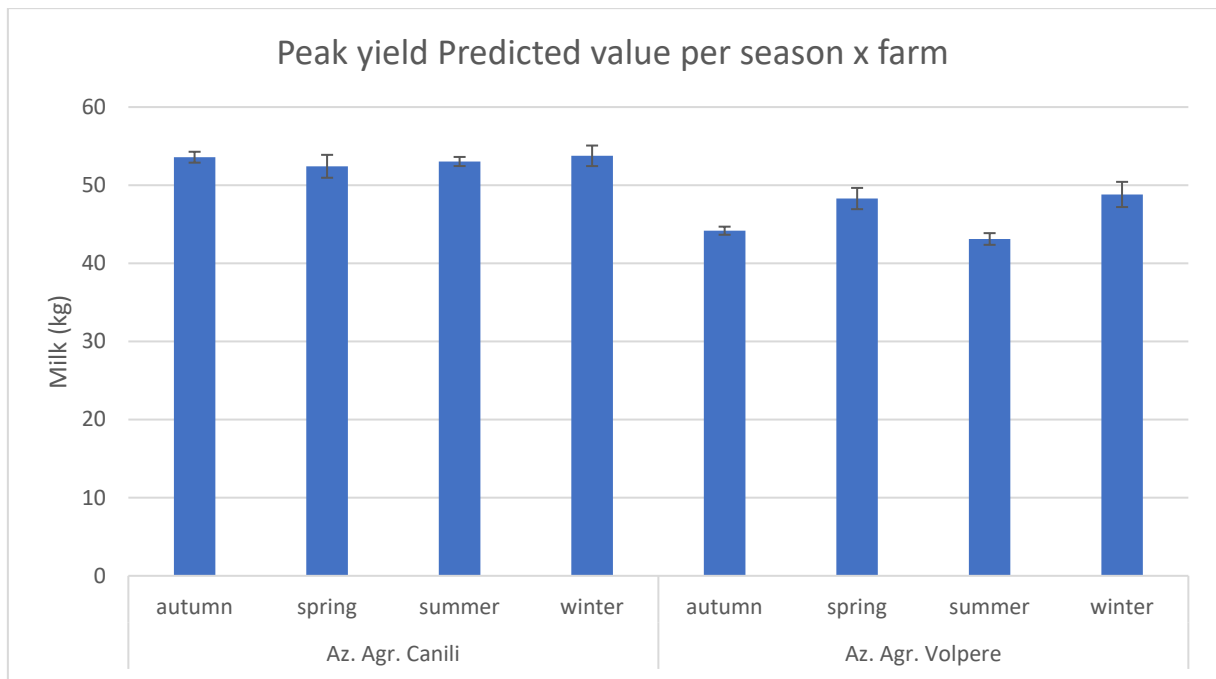
Figure 21: peak yield predicted value per season x farm

GWAS

After imputation of low panels to medium density panels (85 K) and a quality control (QC) for the exclusion of poorly genotyped and faulty data, a total of 78,022 SNPs and 775 cows were kept. Two significant thresholds were used in the GWAS: a Bonferroni threshold (0.05/number of SNPs passing the QC) and a suggestive threshold of -log10(p-value) equals to 4. No SNPs that exceed the Bonferroni threshold were found. The Manhattan plots illustrate the chromosomal distribution of SNPs significantly associated with each trait.

For all three parameters (peak time, peak yield, and lactation persistence for milk in kg), we found suggestive significant SNPs (-log10(p-value) equals to 4) which were spread over several chromosomes. This highlights the complex nature of those traits where several QTLs might have an impact.

For lactation persistence, six SNPs were found suggestive (-log(p-value)=4) which are located one in BTA3, 3 in BTA9, and one in BTA14 and BTA15 (Figure 22).

On BTA3, where the significant SNPs are located, there are several QTLs related to meat and carcass quality, milk composition and body weight.

The areas where the significant SNPs are loated on BTA9 and the area on BTA15 overlapped with QTLs related to fitness aspect such as lifetime production and fertility related traits, but also to quality aspect for instance milk protein percentage and Kappa casein.

On BTA14 the significant SNP overlaps with a QTL area containing QTL related to milk quality, bodyweight and somatic cell score. Overall, the total numbers of genes found for lactation persistence (LP) is 51 across 4 different BTA, highlighting once again the complex and polygenic architecture of this trait (Table 18 and Table 19).

Figure 22: Manhattan plot of the GWAS result for lactation persistence based on imputed whole-genome sequence data. Suggestive SNPs are represented by red dots.

For peak time, three SNPs were found suggestive (-log(p-value)=4). These SNPs are located in BTA6, BTA7, BTA17 and BTA18 (Figure 23).

ON BTA6, where the significant SNPs are located, there is a QTLs region related to clinical mastitis, milk quality, carcass weight. In contrast, on BTA7 the QTLs region is linked to milk composition, somatic cell score, cold tolerance, and milking speed. On BTA17 there is a significant SNP that overlaps with a QTLs area containing QTL related to Blood parameters, milk yield and body weight. Finally, on BTA18 the significant SNPs are located within a QTLs

area coding for milk fat yield, milking speed, milk fat and somatic cell score. Overall, the total numbers of genes found for peak time (PT) is 42 across 4 different BTA (Table 18 and Table 19).
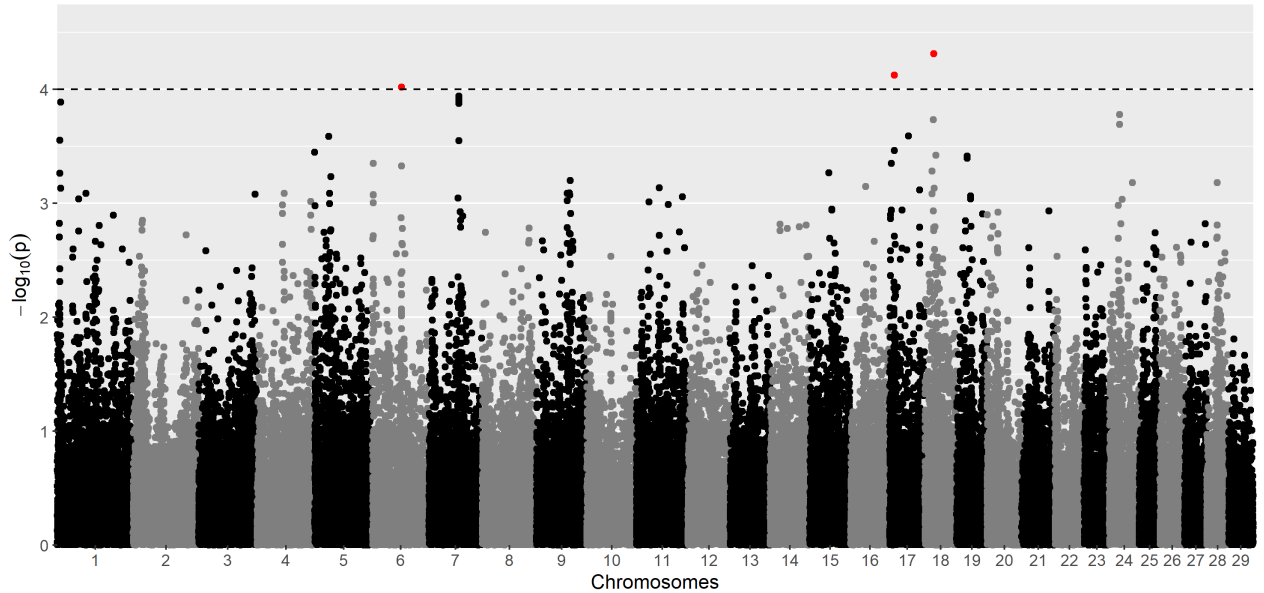
Figure 23: Manhattan plot of the GWAS result for peak time based on imputed whole genome sequence data. Suggestive SNPs are represented by red dots.

For peak yield, four SNPs were found suggestive (-log(p-value)=4). These SNPs are located in BTA1, BTA4, BTA13 and BTA15 (Figure 24).

On BTA1 there is a QTLs region related to udder traits, milk fat yield, pta type and body weight. On BTA4 there is a QTLs region linked to body growth, heifer pregnancy, milk composition and yield. For BTA13 the QTLs region is coding for conception rate, stillbirth, teat length, blood parameters, milk composition and meat and carcass traits.

Lastly, on BTA15 the significant SNP overlaps with a QTL area containing QTL related to semen volume, body height, immuno-capacity, meat and carcass, milk composition, somatic cell score and general reproduction parameters. Overall, the total numbers of genes found for peak yield (PY) is 18 across 4 different BTA as shown in Table 18 and Table 19.
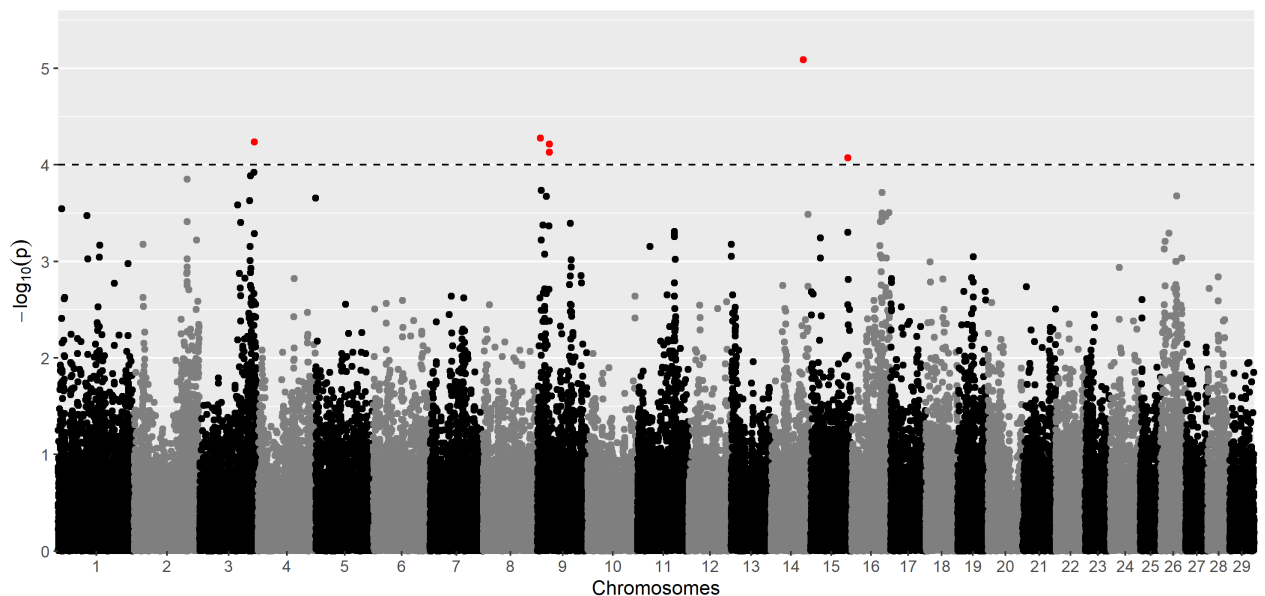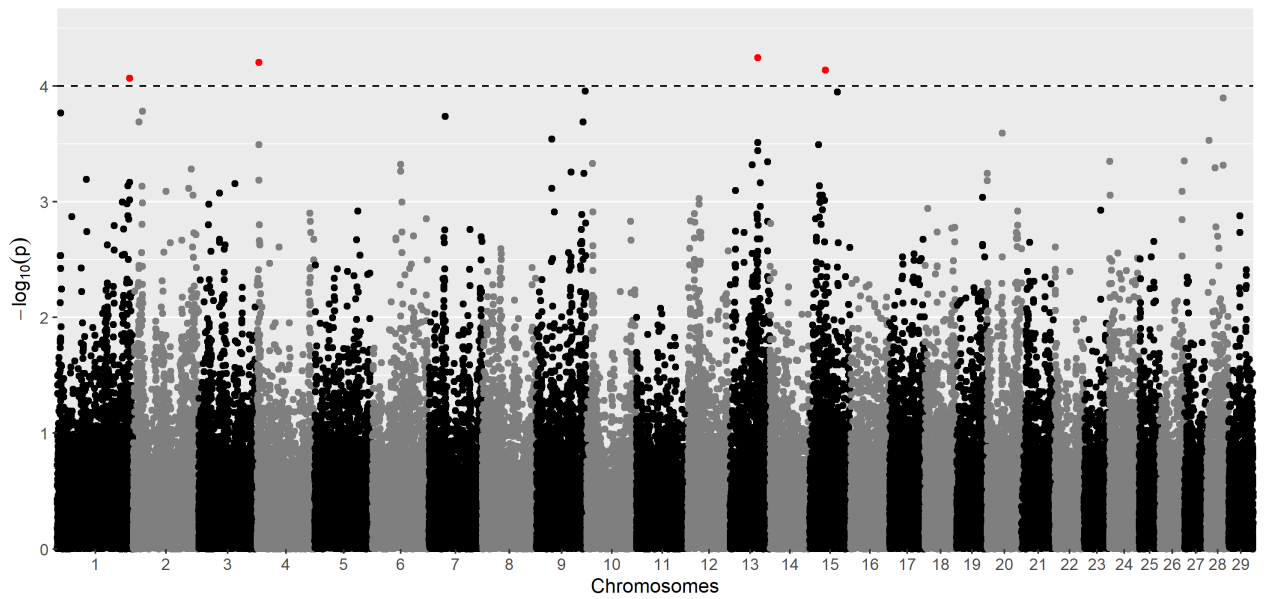
Figure 24: Manhattan plot of the GWAS result for peak yield based on imputed whole genome sequence data. Suggestive SNPs are represented by red dots.

Table 18: Description of SNPs suggestive for Lactation persistence, Peak Time and Peak Yield as well as the QTLs found in those regions.

Trait	SNP	BTA	position	QTLs
peak_time	ARS-BFGL-NGS-106191	18	18905575	Milk fat yield, milking speed, milk FA, somatic cell score
peak_time	ARS-BFGL-NGS-22126	17	10344022	Blood parameters, milk yield and body weight
peak_time	BovineHD0700018717	7	62534252	Milk compositionm somatic cell score and cold tolerance, milking speed
peak_time	BTA-76499-no-rs	6	60369642	Clinical mastits, milk quality, carcass weight
peak_yield	Hapmap49963-BTA-33040	13	57962862	Conception rate, stillbirth, teat length, blood parameters, milk composition and meat and carcass traits
peak_yield	Hapmap52491-rs29011086	4	4229759	Body growth, heifer pregnancy, milk composition and yield

peak_yield	ARS-BFGL-NGS-112072	15	32858415	Semen volume, body height, immuna capacity, meat and carcass, milk composition, somatic cell score, general reproduction parameters
peak_yield	Hapmap53066-rs29026782	1	150635313	Udder traits, milk fat yield, pta type, body weight
persistence	Hapmap29974-BTA-129366	14	67294849	Milk quality, body weight, somatic cell score
persistence	BovineHD0900001298	9	6093741	Length of productive life, non-return rate, inseminations per conceptio,
persistence	BovineHD0300032910	3	113406725	Meat and carcass, milk composition and body weight
persistence	BovineHD0900006728	9	24693972	Milk composition, body weight, averga daily gain, calving ease, milk yield
persistence	BTB-00618478	15	77546541	Lifetime production, milk yield, milk protein percentage, milk kappa casein
persistence	BovineHD0900006739	9	24739623	Average daily gain, body weight, residual feed intake

Table 19: Description of genes found within the ± 50 Mb from the position of significant associated SNPs with Lactation persistence, Peak Time and Peak Yield.

Gene stable ID	Gene name	Gene start (bp)	Gene end (bp)	Chromosome
ENSBTAG00000011001	ERG	150668926	150803833	1
ENSBTAG00000000973	KCNJ15	150591084	150595749	1
ENSBTAG00000037616	KCNJ6	149814196	150141048	1
ENSBTAG00000042581	SNORA72	150366152	150366283	1
ENSBTAG00000045272	U6	150401192	150401298	1
ENSBTAG00000003928	ATG16L1	112967754	113012711	3
ENSBTAG00000020616	DGKD	113062330	113172640	3
ENSBTAG00000018756	DNAJB3	113406610	113407705	3
ENSBTAG00000024726	HJURP	113516126	113528914	3
ENSBTAG00000020173	INPP5D	112794123	112931963	3
ENSBTAG00000017434	MROH2A	113431715	113489004	3
ENSBTAG00000021480	SAG	113023696	113054163	3
ENSBTAG00000044631	SCARNA6	113004991	113005252	3
ENSBTAG00000002030	SPP2	113703201	113730900	3
ENSBTAG00000014652	TRPM8	113586057	113672155	3
ENSBTAG00000026181	UGT1A6	113300772	113425176	3
ENSBTAG00000000149	USP40	113167613	113257470	3
ENSBTAG00000023806	COBL	4582609	4879280	4
ENSBTAG00000042558	7SK	60576749	60577071	6
ENSBTAG00000027569	APBB2	59713923	60095904	6
ENSBTAG00000010677	LIMCH1	60224068	60573254	6
ENSBTAG00000044166	PHOX2B	60646949	60650155	6
ENSBTAG00000020367	SLC30A9	60859705	60946348	6
ENSBTAG00000043958	TMEM33	60814779	60834434	6
ENSBTAG00000005078	UCHL1	60147025	60159287	6
ENSBTAG00000043848	5S_rRNA	62050494	62050611	7
ENSBTAG00000014809	ANXA6	62357582	62404880	7
ENSBTAG00000008340	ATOX1	62942179	62957525	7
ENSBTAG00000053108	bta-mir-11992	62060887	62060946	7

ENSBTAG00000002834	CCDC69	62434491	62469969	7
ENSBTAG00000015625	DCTN4	62073639	62106751	7
ENSBTAG00000003498	FAT2	62700927	62775320	7
ENSBTAG00000020309	G3BP1	62972675	63001442	7
ENSBTAG00000014395	GLRA1	63021619	63108191	7
ENSBTAG00000021829	GM2A	62495947	62508685	7
ENSBTAG00000043553	GPX3	62284773	62292964	7
ENSBTAG00000025136	MYOZ3	62030519	62047672	7
ENSBTAG00000030366	RBM22	62054328	62064514	7
ENSBTAG00000016094	SLC36A1	62635851	62681653	7
ENSBTAG00000020809	SLC36A2	62558340	62584878	7
ENSBTAG00000006624	SLC36A3	62521906	62543906	7
ENSBTAG00000010487	SMIM3	62132825	62152726	7
ENSBTAG00000014835	SPARC	62874446	62897221	7
ENSBTAG00000012671	TNIP1	62293510	62337815	7
ENSBTAG00000043747	U2	62077274	62077461	7
ENSBTAG00000025124	ZNF300	62232854	62240214	7
ENSBTAG00000044579	bta-mir-2284o	25151060	25151120	9
ENSBTAG00000026704	CENPW	24949872	24958580	9
ENSBTAG00000020397	APCDD1L	57899598	57910385	13
ENSBTAG00000051423	bta-mir-6123	57550668	57550749	13
ENSBTAG00000016926	C13H20orf85	58171229	58187773	13
ENSBTAG00000047223	GNAS	57465320	57532635	13
ENSBTAG00000016724	NPEPL1	57665835	57682867	13
ENSBTAG00000018053	RAB22A	57991890	58045639	13
ENSBTAG00000003872	STX16	57697575	57720075	13
ENSBTAG00000017424	VAPB	57923612	57971092	13
ENSBTAG00000030067	bta-mir-584-1	67019700	67019775	14
ENSBTAG00000011908	CPQ	66989157	67551536	14
ENSBTAG00000014357	SDC2	67588233	67708225	14
ENSBTAG00000011572	TSPYL5	66852944	66854263	14
ENSBTAG00000021002	ACP2	77118095	77125767	15
ENSBTAG00000051215	AGBL2	77451766	77481415	15
ENSBTAG00000009574	ARFGAP2	77063167	77075217	15

ENSBTAG00000029881	bta-let-7a-2	32812702	32812774	15
ENSBTAG00000036369	bta-mir-100	32818528	32818603	15
ENSBTAG00000029925	bta-mir-125b-1	32763901	32763988	15
ENSBTAG00000052010	bta-mir-2318	77154599	77154677	15
ENSBTAG00000002520	CELF1	77288832	77368002	15
ENSBTAG00000009573	CSTPP1	76852552	77063224	15
ENSBTAG00000020999	DDB2	77097152	77118221	15
ENSBTAG00000055123	FAM180B	77384965	77387225	15
ENSBTAG00000020911	FNBP4	77492393	77513241	15
ENSBTAG00000018482	KBTBD4	77374325	77379012	15
ENSBTAG00000021700	MADD	77139930	77180691	15
ENSBTAG00000018742	MTCH2	77409614	77429460	15
ENSBTAG00000021707	MYBPC3	77180952	77199310	15
ENSBTAG00000018483	NDUFS3	77379153	77383888	15
ENSBTAG00000010681	NR1H3	77125341	77139490	15
ENSBTAG00000021125	NUP160	77526748	77570322	15
ENSBTAG00000051670	OR4B1	77899334	77900263	15
ENSBTAG00000031025	OR4B1F	77930010	77930939	15
ENSBTAG00000054962	OR4B1GP	78022381	78023309	15
ENSBTAG00000049550	OR4B1H	77914190	77915140	15
ENSBTAG00000031119	OR4S1	77971869	77972819	15
ENSBTAG00000054640	OR4X16	77992130	77993059	15
ENSBTAG00000052223	OR4X17	78006018	78006947	15
ENSBTAG00000053247	OR4X2	77938056	77938985	15
ENSBTAG00000055007	OR4X2B	78028587	78029516	15
ENSBTAG00000031026	OR4X5	77950829	77951758	15
ENSBTAG00000009576	PACSIN3	77074376	77083628	15
ENSBTAG00000021744	PSMC3	77250425	77258901	15
ENSBTAG00000018479	PTPMT1	77368219	77373977	15
ENSBTAG00000024715	PTPRJ	77686573	77861198	15
ENSBTAG00000021745	RAPSN	77261076	77273571	15
ENSBTAG00000019436	SLC39A13	77240232	77248528	15
ENSBTAG00000021709	SPI1	77200844	77230554	15
ENSBTAG00000000842	UBASH3B	33291675	33440714	15

ENSBTAG00000002531	ARHGAP10	10027735	10405085	17
ENSBTAG00000013674	EDNRA	10596771	10669858	17
ENSBTAG00000027182	NR3C2	9583647	10018758	17
ENSBTAG00000010381	PRMT9	10454873	10487503	17
ENSBTAG00000010376	TMEM184C	10488576	10517061	17
ENSBTAG00000006208	ADCY7	18655907	18720625	18
ENSBTAG00000021575	BRD7	18722067	18766993	18
ENSBTAG00000033078	CNEP1R1	18450867	18465539	18
ENSBTAG00000006291	CYLD	19131365	19199820	18
ENSBTAG00000016640	HEATR3	18480139	18519650	18
ENSBTAG00000012295	NKD1	18940443	19034788	18
ENSBTAG00000020936	NOD2	19089956	19119730	18
ENSBTAG00000003961	SNX20	19065512	19074769	18
ENSBTAG00000007942	TENT4B	18573010	18645765	18

Functional analyses of candidate genes

Gene ontology (GO) enrichment analyses were performed to better understand the functional role of the candidate genes identified. GO terms for 7 biological processes were significantly enriched as shown in Table 20.

Table 20: Most significantly enriched gene ontology (GO) terms of candidate genes for Lactation Persistence, Peak Time and Peak Yield.

PANTHER Pathways	Bos taurus REFLIST (23842)	Gene input (102)	Enrichment	Raw p-value	FDR –p-value
Histamine H2 receptor mediated signaling pathway (P04386)	29	3	24.18	3.40E-04	5.48E-02
Beta3 adrenergic receptor signaling pathway (P04379)	33	3	21.25	4.84E-04	3.89E-02

	Bos taurus - REFLIST (23842)	Gene input (102)	Enrichment		
5HT4 type receptor mediated signaling pathway (P04376)	38	3	18.45	7.11E-04	3.82E-02
Enkephalin release (P05913)	44	3	15.94	1.06E-03	3.42E-02
Beta2 adrenergic receptor signaling pathway (P04378)	50	3	14.02	1.51E-03	4.04E-02
Beta1 adrenergic receptor signaling pathway (P04377)	51	3	13.75	1.59E-03	3.66E-02
Endothelin signaling pathway (P00019)	95	4	9.84	8.57E-04	3.45E-02

Bos taurus - REFLIST (23842): number of genes in the Bos taurus reference genome; ***Gene input (102):*** number of gene evaluated as candidate genes from the GWAS analysis; **Enrichment:** evaluation of the GO term enrichment.

Conclusion

The persistence of the lactation curve is a key factor to evaluate to overcome the need of extended lactations that last longer than 305 days and are still profitable.

This latter aspect might have several positive effects since it allows animals to produce more in terms of total production in kg, without implying more production at the peak or at the stage when the animal is undergoing energy deficit.

In this way it's possible to limit metabolic problems, which for a large part, affect the first phase of lactation (Roche et al.,2009).

In addition, a longer lactation allows for a greater window of calving conception, breeding animals beyond the period of negative energy balance allowing for greater fertility and fewer insemination interventions (Allore and Erb.,2000). A calving interval (CI) of 13 months instead of 12 months has been shown to be more economically rewarding (Holmann et al,. 1984).

Furthermore, it is known that peak production is negatively correlated with fertility. Persistence has been shown to be significantly higher in primiparous cows and to be negatively correlated with peak lactation (Sorensen et al., 2008).

In this study, it is observable that the tertiparous category, which are the animals with the highest lactation peak, have the lowest value of lactation persistence.

It is known that persistence is also affected several environmental factors, mainly resulting from farm management which can have a huge impact.

Milking frequency, especially in early lactation, plays a very important role on current and future production throughout lactation (Bernier-Dodier et al,. 2010).

Similarly, the role of feeding influences persistence as seen by (Sorensen et al,. 2008).

This study found that more frequent milking results in an instantaneous 15%-20% increase in production and also better persistence that is maintained for much of the duration of lactation, until the dynamics associated with the last third of pregnancy result in a rapid decrease.

Net of environmental effect and management-related traits, persistence, although it is a highly polygenic character, can be selected since in this study several SNPs associated with this trait were found.

It emerged that the use of imputed whole-genome data for GWAS enabled the identification of a good number of SNPs associated with lactation persistency and milk production traits in dairy cattle.

Several genomic regions and candidate genes were identified for lactation persistency, which are widely distributed across all autosomal chromosomes, especially BTA3, BTA9, BTA14, and BTA15.

For peak time four genomic regions and candidate genes were identified and distributed across autosomal chromosomes specifically on BTA6, BTA7, BTA17 and BTA18.

For what concern peak yield, four regions and candidate genes were found distributed on chromosomes BTA1, BTA4, BTA13 and BTA15.

Numerous candidate genes were found: a total of 111 genes were found for these three traits. Specifically, 51 genes have been found for lactation persistence, 42 genes have been found for peak time, and 18 genes were found for peak yield.

These findings contribute to further understand the molecular mechanisms behind the phenotypic expression of lactation persistency and milk production traits like peak time and peak yield. This result can be useful to improve the genomic evaluation of those economically relevant traits in the Holstein cattle, starting to introduce these traits into genomic indices for the selection of the best individuals.

References

Aken, B.L., S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. Fernandez Banet, K. Billis, C. García Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F.J. Martin, D.N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y.A. Tang, J.-H. Vogel, S. White, A. Zadissa, P. Flicek, and S.M.J. Searle. 2016. The Ensembl gene annotation system. Database 2016:baw093. doi:10.1093/database/baw093.

Allore, H.G., and H.N. Erb. Simulated effects on dairy cattle health of extending the voluntary waiting period with recombinant bovine somatotropin.

Arbel, R., Y. Bigun, E. Ezra, H. Sturman, and D. Hojman. 2001. The effect of extended calving intervals in high-yielding lactating cows on milk production and profitability. J Dairy Sci 84:600–608. doi:10.3168/jds.S0022-0302(01)74513-4.

Bakri, N.E., C. Pieramati, F.M. Sarti, S. Giovanini, and M. 'Naouer Djemali. 2022. Estimates of genetic parameters and genetic trend for Wood's lactation curve traits of Tunisian Holstein–Friesian cows. Trop Anim Health Prod 54:223. doi:10.1007/s11250-022-03219-2.

Bell, A.W. 1995. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. J Anim Sci 73:2804–2819. doi:10.2527/1995.7392804x.

Bernier-Dodier, P., L. Delbecchi, G.F. Wagner, B.G. Talbot, and P. Lacasse. 2010. Effect of milking frequency on lactation persistency and mammary gland remodeling in mid-lactation cows1. J Dairy Sci 93:555–564. doi:10.3168/jds.2009-2320.

Butler, W.R. 2000. Nutritional interactions with reproductive performance in dairy cattle.

Dekkers, J.C.M., J.H. ten Hag, and A. Weersink. 1998. Economic aspects of persistency of lactation in dairy cattle.

Do, D.N., N. Bissonnette, P. Lacasse, F. Miglior, M. Sargolzaei, X. Zhao, and E.M. Ibeagha-Awemu. 2017. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. J Dairy Sci 100:1955–1970. doi:10.3168/jds.2016-11910.

Drackley, J.K. 1999. ADSA foundation scholar award: Biology of dairy cows during the transition period: The final frontier?. J Dairy Sci 82:2259–2273. doi:10.3168/jds.s0022-0302(99)75474-3.

Hale, S.A., A. v. Capuco, and R.A. Erdman. 2003. Milk yield and mammary growth effects due to increased milking frequency during early lactation. J Dairy Sci 86:2061–2071. doi:10.3168/jds.S0022-0302(03)73795-3.

Holden, S.A., and S.T. Butler. 2018. Review: Applications and benefits of sexed semen in dairy and beef herds. Animal 12:s97–s103. doi:10.1017/S1751731118000721.

Holmann, F.J., C.R. Shumway, R.W. Blake, R.B. Schwart, and E.M. Sudweeks. 1984. Economic Value of Days Open for Holstein Cows of Alternative Milk Yields with Varying Calving Intervals. J Dairy Sci 67:636–643. doi:10.3168/jds.S0022-0302(84)81349-1.

Hu, Z.-L., C.A. Park, and J.M. Reecy. 2022. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res 50:D956–D961. doi:10.1093/nar/gkab1116.

Larmer, S.G., M. Sargolzaei, L.F. Brito, R. v. Ventura, and F.S. Schenkel. 2017. Novel methods for genotype imputation to whole-genome sequence and a simple linear model to predict imputation accuracy. BMC Genet 18. doi:10.1186/s12863-017-0588-1.

Mallard, B.A., J.C. Dekkers, M.J. Ireland, K.E. Leslie, S. Sharif, C.L. Vankampen, L. Wagter, and B.N. Wilkie. 1998. Alteration in Immune Responsiveness during the Peripartum Period and Its Ramification on Dairy Cow and Calf Health. Pages 585–595 in Journal of Dairy Science. American Dairy Science Association.

Österman, S., and J. Bertilsson. 2003. Extended calving interval in combination with milking two or three times per day: Effects on milk production and milk composition. Livest Prod Sci 82:139–149. doi:10.1016/S0301-6226(03)00036-8.

Pedrosa, V.B., F.S. Schenkel, S.Y. Chen, H.R. Oliveira, T.M. Casey, M.G. Melka, and L.F. Brito. 2021. Genomewide association analyses of lactation persistency and milk production traits in holstein cattle based on imputed whole-genome sequence data. Genes (Basel) 12. doi:10.3390/genes12111830.

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing.

Rajala-Schultz, P.J., J.S. Hogan, and K.L. Smith. 2005. Short communication: Association between milk yield at dry-off and probability of intramammary infections at calving. J Dairy Sci 88:577–579. doi:10.3168/jds.S0022-0302(05)72720-X.

Roche, J.R., N.C. Friggens, J.K. Kay, M.W. Fisher, K.J. Stafford, and D.P. Berry. 2009. Body condition score and its association with dairy cow productivity, health, and welfare. J Dairy Sci 92:5769–5801. doi:10.3168/jds.2009-2431.

Smith, K.L., D.A. Todhunter, and P.S. Schoenberger. 1985. Environmental Mastitis: Cause, Prevalence, Prevention. J Dairy Sci 68:1531–1553. doi:10.3168/jds.S0022-0302(85)80993-0.

Sorensen, A., D.D. Muir, and C.H. Knight. 2008. Extended lactation in dairy cows: Effects of milking frequency, calving season and nutrition on lactation persistency and milk quality. Journal of Dairy Research 75:90–97. doi:10.1017/S0022029907002944.

Tekerli, M., Z. Akinci, I. Dogan, and A. Akcan. 2000. Factors affecting the shape of lactation curves of Holstein cows from the Balikesir Province of Turkey. J Dairy Sci 83:1381–1386. doi:10.3168/jds.S0022-0302(00)75006-5.

WOOD, P.D.P. 1968. Factors affecting Persistency of Lactation in Cattle. Nature 218:894–894. doi:10.1038/218894a0.

Appendix

Appendix I

Farm	
ASL code	160BS004
Milk destination	Dietary milk (qualified for Grana Padano)

Organization chart	
Family management	Yes
Employees	Yes(X) No
Employees (n°)	1

Farm buildings	
Year of construction:	
1969 first barn	
2019 actual barn with 150 berths	
2019 milking parlor	
Animals	
Presence of other races:	Yes/ Jersey
Other races (n° animals)	3
Lactating cows:	143
Total number of bovines:	380

Stalling			
Group:	Stall type:	Floor:	Bedding:
Lactating cows	Berths	Concrete floor	Blister/ Pelleted straw
Dry cows	Permanent bedding	Concrete floor	Straw
Infirmary/Birthing room	Permanent bedding	Concrete floor	Straw
Heifers	Permanent bedding	Concrete floor	Straw
Calves	Permanent bedding	Concrete floor	Straw

Pasture	
Group:	Days (n°):
Dry cows	Only from May to September (60 days)

Ventilation and cooling	
Nebulizers	Yes, with fans
Showers	Yes

fans	Yes
Destratifiers	Yes

Milking parlor	
Type:	Herringbone
Number of milking stall:	10 + 10
Year of construction:	2019

Feeding:	
TMR:	Yes
Autofeeders:	Yes (Calves)
Daily administration:	Groups:
2	Fresh cows
2	Late lactation cows
1	Heifers
1	Dry cows
1	Calves

TMR:	
Feed:	Kg:
Chopped straw	400 g
Corn meal	5 kg
Soybean	5,5 kg
Mineral supplementation	1 kg
Wheat sillage	5 kg
Whole earlage	3 kg
Corn sillage	35 kg

Appendix II

Farm	
Code ASL	007MN10
Milk destination	Dietary milk (Granlatte cooperative)

Organization chart	
Family management	Yes
Employees	Yes(X) No
Employees (n°)	11

Farm buildings	
Year of construction:	
1996 first barn	
2003 Second barn	
2011 enlargement lactating cows barns	
2015 heifers barn	
2015 dry cows barn	
2020 lactating cows barn	
2020 rotary milking parlor	
Animals	
Presence of other races:	Yes/Jersey
Other races (n° animals)	6
Lactating cows:	720
Total number of bovines:	1500

Stalling			
Group:	Stall type:	Floor:	Bedding:
Milking cows	Berths	Concrete floor	Blister/sand/ unchopped straw
Dry cows	Berths + paddock	Concrete floor	Blister/sand/unchopped straw
Infirmary/Birthing room	Permanent bedding	Concrete floor	Straw
Heifers	Permanent bedding	Concrete floor	Straw
Calves	Permanent bedding	Concrete floor	Straw

Pasture	
Group:	Days (n°):
Dry cows	45

Ventilation and cooling	
Nebulizers	No
Showers	Yes
fans	Yes
Destratifiers	Yes

Milking parlor	
Type:	Rotary
Number of milking stall:	60
Year of construction:	2020 (they turned to rotary on 29/04/21 but they milked cows with both plants for a week)

Feeding:	
TMR:	Yes
Autofeeders:	Yes (for calves, with milk)
Daily administration:	Groups:
3	Fresh cows
2	Late lactation cows
1	Dry cows
1	Heifers
3-4 times per week	Calves

TMR	
Feed:	Kg:*
Supplement	
Soybean/corn mix	
Protein mix	
Soybean	
Corn meal	
Molasses	
Cotton seed	
Wheat hay	
Alfa alfa hay	
Alfa alfa sillage	
Earlage	
Wheat sillage	
Corn sillage	
Brewers	

*Quantities are subjected to change depending on availability and season.