UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO IN SANITÀ E PRODUZIONI ANIMALI: SCIENZA, TECNOLOGIA E BIOTECNOLOGIE

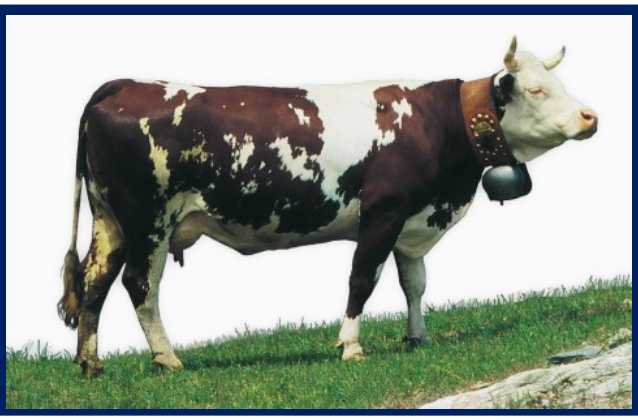
DOTTORATO DI RICERCA IN PRODUZIONI ANIMALI

XXIV CICLO

"Genomic aspects of genetic improvement for mastitis resistance in dairy cattle"

Sergio Iván ROMÁN-PONCE

Docente guida: Prof. Alessandro BAGNATO.
Correlatore: Dr. ssa Antonia Bianca SAMORÉ.



Anno accademico 2010/2011

To god, no matter the name,

To my mother†, you are always with me,

To my wife, the reason of my life,

To my family, thanks don´t let me down,

To my friends, for your advices,

Genomic aspects of genetic improvement for mastitis resistance in dairy cattle

by

Sergio Ivan Roman Ponce

A dissertation submitted to the Graduate School of Animal production of the

Universitá degli Studi di Milano to partial fulfillment of the requirements for the degree of Doctor in Animal Production.

Milan, Italy

2012

Approved by:

Prof Alessandro Bagnato	Prof. G. Matteo Crovetto

BIOGRAPHY

Sergio Iván Román-Ponce was born on September 5th, 1977 in Veracruz, Mexico. His origins come from the North of the state of Veracruz, Mexico. He was raised in Tres Valles, Veracruz, Mexico. Sergio's vacations were mostly in the grandparents's farm of dual purpose crossbreed cattle. This stimulates him to hire the Faculty of Veterinarian at the Universidad Veracruzana (State University of Veracruz).

He graduated as third in the class 2000 as Doctor in Veterinarian. During his first year of professional life, he developed the interest in animal breeding and genetics. As results he attended the Master of Science degree in the Faculty of Veterinarian at the Universidad Nacional Autonoma de Mexico (National Autonomous University of Mexico). One year after his competition he started to work in the Instituto Nacional de Investigaciones Forestales, Agricolas y Pecuarias (National Research Institute for Forest, Agricultural and Livestock) since October 1st, 2007.

To complete his academic training, he get enroll to obtain the degree of Doctorate in Animal Production in the graduate school at the University of Milan. He plans to return to work in the Instituto Nacional de Investigaciones Forestales, Agricolas y Pecuarias in Mexico.

ABSTRACT

ROMÁN-PONCE, SERGIO IVÁN. **Genomic aspects for genetic improvement of mastitis resistance in dairy cattle** (Under the supervision of Alessandro Bagnato).

The overall objective of this work is to evaluate the mastitis resistances genetic aspects in cattle with genetic and genomic tools. This was accomplished by: 1) the estimation of genetic parameters for traits related to udder health in the Valdostana cattle breed; 2) the evaluation of the effect on the genomic breeding value estimation of mastitis traits of the assumption of different prior probability values for the proportion of markers with a large effect; 3) the estimation of the fraction of the genetic variance not explained by the 54K Illumina SNP chip while using different marker-based relationship matrices; and 4) the exploration of the influence of the level of phenotype accuracy on the genomic predictions, by including phenotypes with different minimum level of reliability in the training population to estimate the marker effects. For the objective one, data of the Valdostana cattle breed were used for a total of 34,291 milking cows, collected born from 2002, with the milk bacteriological analysis that reported the presence of *Staphylococcus aureus*, *Streptococcus agalactie*, *Staphylococcus* ssp., *Streptococcus* ssp., *Escherichia coli*, minor pathogens gram positive, minor pathogens gram negative and fungi. Data used also included information on SCS and milk yield (MY) and genealogical data were extracted from the national herd book. The threshold model analysis was evaluated to be the most appropriate approach for the binary data under analysis concerning the presence/absence of pathogens in milk. Generally moderate heritability values (from 0.02 to 0.09) were estimated for the specific presence of the pathogens. This suggested that bacteriological data can be considered for the genetic selection to improve udder health. Promising results in mastitis selection were predicted through the aggregation of the indirect indicator (SCS), the trait commonly worldwide used in dairy cattle mastitis selection, and innate resistance to some of the major pathogens causing mastitis, such as *Staphylococcus aureus, Streptococcus agalactie and Escherichia coli.* The Bovine SNP50 Illumina genotypes of 1,089 Brown Swiss bulls were used for the second purpose. A total of 51,582 SNP markers were considered in the analyses after the exclusion of markers on chromosome X. The estimated breeding values of bulls were provided by the Italian Brown Cattle Breeders Association for SCS and for the following production traits: milk yield, fat yield, protein yield, fat percentage, and protein percentage. To perform genomic breeding value estimation,

data available were split in two populations subsets: training (the 846 bulls born before 2001) and test population (243 bulls born from 2001 and 2005). The value assumed (0.001, 0.005, 0.01, 0.05, 0.1 and 0.5) for the number of SNP with a large effect did not have impact on the marker effect estimates or on genomic predictions accuracies. For the objective three, genotypes from Bovine SNP50 chip of 1,086 sires were available for a total of 35,706 SNP with a 99.34% total genotyping rate. Phenotypic information consisted on EBV estimated by the National Breeder Association of Brown Swiss dairy cattle for SCS, production and type traits as follows: milk yield, fat yield, protein yield, SCS, overall conformation, stature, and rear leg side view, fore udder attachment, rear udder width, udder support, udder depth, feet and legs and foot height. Three generations of genealogical information were used (4,988 animals). The proportion of the genetic variance addressed by markers was estimated using different marker-based relationship matrices. In all traits considered the fraction of the genetic variance not explained by the genetic markers did not significantly differ from 0 for all the designs including in the training population bulls with high accuracy. Indeed no substantial differences were found with the use of different genomic relationship matrices. The only exception was the genomic matrix corrected by the heterozygosity per SNP. All the analysis with that genomic matrix converged and the genetic variance explained was bigger than with the other matrices. For the objective 4; the phenotype was associated with the genotype at 35,546 SNP markers for 1357 sires for all available traits and genomic EBV estimated using different strategy to identify the training population. Results indicate that selecting the training population with accurate phenotypes yield genomic EBV with larger accuracy. The genetic selection of a complex trait, as mastitis selection, involved several aspects and it would be better searched through the integration of classical genetic aspects and of the genomic selection. The latter genomic approach would benefit of the use of all the aspect considered in this thesis. In particular suggestions were drawn for the use of best prior parameters, the definition of the most informative training population, the calculation of genetic parameters for binary and non-binary traits, and the calculation of genomic relationship variances to be used in genomic breeding value estimations.

TABLE OF CONTENTS

Chapter 1

Literature review

From phenotypic to genomic selection.

Quantitative genetics or genetics of complex traits it is based on models in which many genes act and interact to determinate the traits and in which non-genetic factors (environmental) may also play an important role in phenotype determination (Fisher, 1918; Wright, 1922). The analysis of variance and regression models represent the traditional methods to dissect the genetic variance in terms of additive genetic, interactions of effects between alleles within loci (dominance) and among loci (epistasis) (Falconer and Mackay, 1996; Lynch and Walsh, 1998).

The variance component estimation made possible the calculation of genetic parameters, such as heritability ($\mathbf{h^2}$), the ratio of the additive genetic variance ($\mathbf{V_A}$) and the overall phenotypic variance ($\mathbf{V_P}$) (Falconer and Mackay, 1996). The genetic parameters are basic information to predict the breeding values of individuals (expected performance of the offspring) and the genetic response in the genetic improvements programs (Hazel, 1943).

In the middle of the last century Henderson (1984) proposed the use of mixed models for the variance component estimation. The mixed models equations enable the simultaneous incorporation of fixed (usually environmental ones) effects and random (usually the additive genetic component) effects (Lynch and Walsh, 1998; Sorensen and Gianola, 2002). The most common mixed models applied in animal breeding have been the sire model and the animal model to predict the additive genetic value. Predictions obtained with these models are defined best linear unbiased prediction or **BLUP** (Henderson, 1953; Henderson, 1984).

In the last three decades, the worldwide efforts on genetic improvements were concentrated into the conventional progeny test aimed to predict the breeding value of sires. The objective has been to find the group of sires interesting for the seedstock industry. In general, the conventional scheme takes on average 54 months and a big amount of economic resources to prove the bulls (Schaeffer, 2006). However, the implementation of progeny test schemes and genetic evaluations occurred over the entire worldwide animal breeding industry.

The continuous exchanges of sires among the countries, the differences in genetic evaluation models, breeding objectives, genetic level and farming environments

increased the complexity of the comparison between sires from different countries. For that reason, in 1983, the International Committee for Animal Recording, the European Association for Animal Production and the International Dairy Federation established Interbull. Since 1996, the European Union appointed the Interbull Centre as the community reference body for international sire evaluations.

More recently molecular genetic technologies allow new approaches to be investigated to improve the efficiency of animal selection. Very recently the genomic analyses of animals DNA and their use in genetic selection is having a strong impact in dairy cattle selection and its industry organization. The knowledge of genomic information on DNA makes possible the evaluation of the genetic merit without the need of progeny performance knowledge and therefore without the need of waiting for the necessary time. Genomic selection is based on the estimation of marker additive effects in a group of animals that have information both on phenotypes and on genotypes. Summing all the marker effects for each animal the "genomic" estimated breeding value is obtained. This estimation is therefore possible both for the group of animals used for the estimation of marker effects (training population: progeny tested bulls) and for other animals (test population: young bulls) that only have genotypes but not phenotypes (Meuwissen *et al*., 2001).

The availability of genomic maps with a high density of multi-locus single nucleotide polymorphism (**SNP**) made possible the introduction of the genomic selection in real populations. The SNP are a DNA sequence variation occurring when a single nucleotide is changed in the genome (Matukumalli *et al*., 2009). The advantage of the application of the genomic selection (**GS**) into the progeny test schemes could be the reduction of the generational interval and the increment of the rate of genetic gain in the breeding programs.

Up today, the base of the animal breeding industry has been the conventional progeny test, which uses direct or indirect measure of the performance of the animals. Now a day with the GS, it is possible to predict the genetic merit of young animals without own performance with sufficient accuracy. This possibility makes possible to incorporate the young sires into a progeny testing scheme after the prediction of the breeding values obtained by markers. The possibility to obtain the genomic values for traits that are difficult or expensive to measure in the entire population, it is an additional advantage.

New information new possibilities (SNP genotyping= new information; genomic approach = new possibility).

The marker-assisted selection (**MAS**) consisted in the selection of animals based on the genetic variances captured by identified quantitative trait loci (**QTL).** This methodology identifies the QTL by the application of the linkage maps as proposed by Beckmann and Soller (1983). The experiences of integration of the identification of QTL results into the genetic improvement programs has been limited because of the small response to selection and the small proportion of genetic variances explained (Dekkers, 2004).

The approach of the QTL discovering improved its efficiency with the availability, at affordable costs, of dense genomic maps of SNP information. The SNP is a binary marker and by itself is less informative than the other molecular markers available. The presence of the non-random association of alleles at two or more loci, named as linkage disequilibrium (**LD**) allows the direct association of single markers to the quantitative trait commonly, commonly referred as genome wide association (**GWAS**).

The principal source of SNP data had been the public databases of SNP, such as bovine genome sequencing efforts at the Baylor College of Medicine and Bovine HapMap project efforts (Matukumalli *et al*., 2009). The large amounts of confirmed SNP in the livestock populations allowed the development of high density SNP arrays.

With the commercial availability of the high density SNP assays in cattle (Illumina SNP chip Bovine HD) one can expect to be able to address a substantial proportion of the genetic variation in complex traits. Through the use of simulated data, Meuwissen et al. (2001) were able to predict the breeding values using markers data alone (less dense than the actual SNP chips) with accuracies up to 0.85.

The theoretical demonstration of the calculation feasibilities (Meuwissen et al., 2001) and the great advantages resulting from GS implementation (Schaeffer, 2006), together with the continuous release of new high dense genotyping arrays at reduced costs (Matukumalli *et al*., 2009), was reflected into the intensification of the genotyping efforts in the dairy cattle populations (Berry *et al*., 2009; Schenkel *et al*., 2009; VanRaden *et al*., 2009).

The principles of genomic selection.

The methodology foundations of the genomic selection were presented by Meuwissen *et al*. (2001). Prediction equations, i.e. the additive marker effects, are calculated in the reference (training) population, using both genotype (dense SNP genotypes) and phenotype data (EBV of proven individuals): The prediction equations can then be used to calculate the genetic values of all individuals of the population, in particular young bulls and bull dams the one mostly contributing to genetic improvement in populations. The best animals will be therefore selected from the whole population according to the estimated breeding values. The main assumptions of genomic selection are: 1) the chromosome segments are the same in the whole population because the markers are in LD with QTL that they bracket; 2) the marker density is sufficient to ensure that all QTL are in LD with a marker or with an haplotype and 3) the proportion of the additive genetic variance explained by the markers is assumed to be close to 100% (Meuwissen *et al*., 2001).

Generally the genomic selection process is performed with the following steps (VanRaden *et al*., 2009):

1) The estimation of SNP effects on genotyped animals using their phenotypic performance, daughter yield deviation (**DYD**) or estimated breeding values (**EBV**).

2) The calculation of direct genomic values (**DGV**) for the selection candidates using the genotype of each animal and the SNP effects estimated in the training population.

3) The calculation of an aggregated genomic EBV (**GEBV)** combining DGV and traditional EBV.

The first measures of the accuracy of the genomic predictions were calculated as the correlations of DGV with the phenotypes, such as DYD, EBV or de-regressed proofs. This methodology has been used to tune the models for the estimation of marker effects. In general it is known that the accuracy on the genomic predictions is influenced by the following parameters: the number of animals with both genotyping and phenotyping information, (Meuwissen *et al*., 2001; Van Raden *et al*., 2009), the additive relationships between the reference and training populations (Legarra *et al*., 2008; Meuwissen, 2009), the density of the marker assay (Meuwissen *et al*., 2001; Calus *et*

al., 2008; Solberg *et al*., 2008), and the heritability of the phenotypes (Calus, 2010; Solberg, 2008).

The possibility to work on new traits.

The genomic selection methodology opens the possibility of considering in selection also traits difficult or expensive to measure that were difficult to select with the classical approach. The main advantage of GS is actually that the marker effects, to estimate DGV, may be calculated from a small group of animals with both phenotypic and genotypic information and that these effects may be then used to estimate DGV of the whole population, also if the trait is not widely recorded.

Feed intake and residual feed intake are two of these traits that can play an important role in the economic efficiency of livestock industry (Rolfe *et al*., 2011). Recently, these traits have been successfully included into some national cattle genetic evaluation system for beef cattle (MacNeil *et al*., 2011) with the prediction of genomic breeding values (Mujibi et al., 2011).

In dairy cattle, the proportion of specific fatty acids in milk (Soyeurt and Gengler, 2008) and the coagulation properties of milk (Cecchinato, *et al*., 2011) are traits that can strongly benefit of the GS implementation. Recently, some promising results of GWA have been conducted for these traits opening possibilities also for the genomic selection for fatty acid milk composition (Bouwman et al., 2011).

The worldwide warming for the climate changes represents a new issue of the livestock production. Livestock produces close to 18% of the worldwide emissions of greenhouse gas as methane, nitrous oxide, water vapor, carbon dioxide, and ozone (Moran and Wall, 2011). Nowadays it is possible to record data on individual gas emission in dairy cows. The availability of individual records on gas production makes possible the estimation of the additive genetic variances (de Haas *et al*., 2011) and thus the selection of individuals for reduced gas emission.

The small number of animals recorded represents the principal limit for the inclusion of these traits in classical genetic selection programs. With the genomic approach, the inclusion of these traits in genetic selection programs will be possible because not all

the animals in the population must have records for those traits: once the SNP effects are estimated, those effects could be used for genomic prediction in the entire population.

Other markers may enter the picture of genomic selection.

The sequencing of the whole genome provides a wealth of information to tackle genetic problems, such as the identification of the molecular basis of complex traits, which are difficult to study with conventional approaches (Hobert, 2010). Furthermore, the structural variations take many forms in the genome, i.e. deletions, insertions, duplications and complex multi-site variants that are collectively named copy number variations (**CNV**). A CNV can be simple in structure, larger than 1kb, such as tandem duplication, or may involve complex gains or losses of homologous sequences at multiple sites in the genome (Rendon *et al*., 2006).

The array comparative genomic hybridization (**aCGH**) has been so far the most used technique to disclose CNV (to detect, validate and characterize). In aCGH experiments genomic DNA samples are co-hybridized on the same oligonucleotide array and the genomic variation differences from the reference sample lead to CNV detection (Shinawi and Cheung, 2008). Studies on CNV identification are available for cattle (Fadista *et al*., 2010; Liu *et al*., 2010), chicken (Wang *et al*., 2010), swine (Fadista *et al*., 2008) and goat (Fontanesi *et al*., 2010).

The high dense marker genotyping platform presents a median gap spacing among markers minor than 3 kb; which make feasible the identification of CNV (Wang *et al*., 2007). Even if CNV identification are affected by the algorithms implemented (Tsuang *et al*., 2010), some studies with high dense SNP arrays have been performed to detect CNV in cattle (Bae *et al*., 2010; Hou *et al*., 2011) and swine (Ramayo-Caldas *et al*., 2010).

The CNV represent a significant source of genetic diversity in mammals covering ~12% of the genome (Rendon *et al*., 2006), and they have been shown to be associated with phenotypes (diseases/traits) in humans (Stankiewicz and Lupski, 2010). According to literature consulted to today, there are not available association studies between CNV and complex traits in dairy cattle. This research topic represents an important source of

information to integrate the CNV in the genomic selection methodologies. The availability of this new source of information is expected to improve the accuracy of genomic breeding values.

The new possibility for the dairy cattle selection industry.

The advantages deriving from the introduction of the genomic selection in the dairy cattle industry have been already evidenced by both deterministic (Schaeffer, 2006) and stochastic (Buch *et al*., 2011) simulations. Recently, some hypothetical scenarios were evaluated comparing situations with or without the incorporation of genomic information. The "turbo" scenario in which young males and females are selected based on parent average GEBV is one of the possibilities**.** In this case the sires are selected based on GEBV and these animals are called genomic bulls and they enter in breeding programs at around 18-21 months, as soon as they are sexually mature (Buch *et al*., 2011).

In this scenario the annual genetic gain was higher than in the conventional progeny test schemes both for milk production (65%) and functional traits (173%). This depended mainly by the smaller generational interval in the turbo scheme than in the conventional progeny testing scheme. The rate of inbreeding predicted was only 0.74% points per generation, i.e. within the international recommendations (de Roos *et al*., 2001).

The possibility to select candidates of sires on the genetic merit that includes the direct genomic breeding value modifies the entire structure of the animal breeding industry dramatically reducing costs (Schaeffer, 2006).

What is so far the genomic application in dairy cattle? – Need of large numbers for training populations. Eurogenomics, **Intergenomics and other.**

In early 2008, the United States of America and Canada joined their forces to create a large training population for Holstein-Friesian dairy cows. The result was the first genomics predictions, which showed the necessity to boost the genomic reliabilities through the aggregation of genotypes across countries (VanRaden *et al*., 2009).

Recently, the United Kingdom and Italy entered into the collaborative genomic evaluation consortium project of North American countries.

In general, the collaborations are originated because of the necessity to increase the reference population, to avoid repeated genotyping, to integrate small breeds and to share algorithms and software. Such as the consortium Eurogenomics integrated by five European national Holstein cattle breeding companies. Each partner participates with at least 4,000 proven bulls to improve the reliability of the genomic breeding values. In total the Eurogenomics reference population is extended to at least 16,000 bulls. The results of these collaborative experiences are increments of 10% in the reliabilities, compared with the reliabilities obtained in the national genomic evaluation with the single national reference populations alone (David *et al*., 2010). These results are different among countries and between traits, but the improvements in the accuracy ranged from 2% to 19% (Lund *et al*., 2011).

Another example of a worldwide collaboration in the genomics fields is the project Intergenomics, which was funded by the European National Associations of Brown Swiss together with the Interbull consortium. The main objective is to develop a common genomic evaluation program for the Brown Swiss guided by Interbull. In that program, the genomic reliabilities of the young bulls were about 2.5 times larger than those of their respective parent averages. Correlations between conventional proofs, DGV and GEBV were close to one (Zumbach *et al*., 2010).

Genomic predictions models.

The statistical foundation of the genomic selection consisted on three methods proposed by Meuwissen *et al*. (2001). These methods were developed to overcome the lack of degrees of freedom when all the markers effects are estimated simultaneously. The size of the training population generally is smaller than the number of markers. The models proposed were:

1) Least Squares (**LS**). Here chromosome segment effects were treated as fixed effects. No assumptions are made regarding the distribution of the effects. All the genes are tested one by one for their statistical significance. Some markers with non-significant effect were close to zero. The least squares approach presented two major inconvenient.

The first is the choice of the significance level and the second is that the chromosome segment effects were estimated from single segment regression.

2) Best Linear Unbiased Prediction (**BLUP**). The allelic effects are fitted as random effects instead of as fixed effects. This helps to avoid the problems with the degree of freedom. All allelic effects could be estimated simultaneously. The assumption is that every marker or chromosome segment effect has the same variance but this was considered in some case not to be appropriate. Normally, the majority of the genes have very little effect on the trait. These genes will be dominant on the estimation of the variance of the allelic effects, i.e. this estimate will be close to zero. With BLUP the chromosome segment (or QTL) with the largest variance tend to have a variance over-estimated, and this will decrease the efficiency of genomic selection.

3) Bayesian estimation (**Bayes**). In theory, this is a method that allows the variance of the chromosome segment effects to vary among segments. This approach represents the opportunity to obtain better estimation of breeding values. Different variance for every gene is assumed to exist and it is estimated based on a given prior distribution. In this way it is possible to take into account the prior knowledge about the distribution of the effects. Some of the chromosome segments with a large additive effect probably contain a QTL while segments with no QTL will have a moderate to small effect.

The prior distribution of the variance for the gene i (V_{ai}) is assumed here:

$V_{ai} = 0$ with probability π

$V_{ai} \sim \chi^2(v,S)$ with probability $(1-\pi)$

where $\boldsymbol{\pi}$ depends on the mutation rate at the gene, and the distribution assumed is the inverse – chi squared distribution with v degrees of freedom and scale parameter S ($\chi^2(v,S)$).

All the three models allow the estimation of prediction equations for the estimation of the genomic breeding values. The estimation of the DGV is obtained summing the markers effect.

The deterministic predictions of the accuracy of the DGV may be used to establish the size of the reference population and to anticipate the level of accuracy in the selection of candidates. Some authors have used the inverse of the mixed model equations to

estimate the accuracy of the DGV (VanRaden, 2008; Hayes *et al*., 2009a,c) but this may overestimate the accuracy as demonstrated by Goddard *et al*. (2011).
Recently, a proposal to estimate the reliability (squared accuracy) of the DGV for each animal in the test population was proposed by Goddard *et al*. (2011). To obtain the accuracies values, two quantities are needed: the proportion of the genetic variance explained by markers and the accuracy with which the combined markers effects are estimated (Dekkers, 2007; Goddard, 2009; Goddard *et al*., 2011).

Design of the training and the testing populations for genomic selection.

The genomic predictions offer exciting opportunities for genetic improvements of livestock. One of them is the wider opportunity to choose younger bulls in the progeny testing structure (Köning *et al*., 2009), which represents both a challenge and a challenge of developing genomic breeding value estimations with the smallest bias and the highest accuracy. Some strategy to optimally organize the populations for the genomic programs have been suggested but the most common design is the use of data sets split into training and testing populations, which avoid the generational overlapping between both data set (Amer and Banos, 2010).

The prediction of breeding values of unrelated individuals is required for some of the most promising application of genomic selection. This is actually the case when it is necessary to use field data coming from elite breeding stocks (Wray *et al*., 2007). In this situation, if it necessary to obtain accuracies of around 0.90, it is necessary to use 10*Ne*L SNPs and 2*Ne*L records in the training data set, where **N** is the effective population size and **L** the genome size in Morgan (Meuwissen, 2009).

The availability of denser maps of markers did not only yield more accurate predictions of the genetic merit, but also enlarges the generation intervals necessary to re-estimate the markers effects. Simultaneously, the size of the training dataset contributes to decrease the differences between the predictions models (G-BLUP and BayesB).

The inclusion of females genotypes in the training population was evaluated by Hugh *et al*. (2011). Their inclusion yields increments in accuracy and genetic gain if compared to traditional BLUP breeding programs. Their studies show alto that generation interval was reduced and inbreeding remained in a reasonable improvement rate.

In general, there are not clear indications on how to integrate the training population to estimate the markers effects to predict the genomic breeding values. In general, the dairy cattle populations present a complex structure of the pedigree, which increase the relationship between both populations, and the possibilities to predict more accurate genomic breeding values across much more generations.

Prior probabilities (π).

The Bayesian methods proposed by Meuwissen *et al*. (2001), Bayes A and B, require the assumption of specific hyperparameters. It has been demonstrated that the values assumed can affect the accuracy of the genomic predictions (Gianola *et al*., 2009). The prior distribution of QTL effects is one of the hyperparameters that concern the accuracy of DGV estimation. It resulted that different assumptions of prior distributions produced DGV highly correlated (Verbyla *et al*., 2010). This seems to suggest that the models are not affected by the choice of prior distributions. Even though, it is recommended that any information about a trait's QTL effect distribution and phenotypic data should be integrated to determine the assumptions regarding the hyper parameters in the prediction of model.

Relationship matrix based on markers information.

Identity by descent (**IBD**) refers to alleles that descend from a common ancestor in a base population (Wright, 1922). This approach leads to the classic estimation of the relationship matrix based on pedigree. This matrix is fundamental to estimate the genetic parameters such as heritability for complex traits (Henderson, 1976). The numerator relationship matrix based on pedigree data goes back to a base population, which is considered unrelated, unselected and non-inbred (Henderson, 1976; Quass, 1976).

The genome-wide genetic markers can capture the additive relationship through the estimation of the relationships matrix (Fernando 1998; Habier *et al.,* 2007; VanRaden, 2008) based on markers, named Gmatrix (**G**). The values obtained correspond to twice the coefficient of coancestry of Malécot (Malécot, 1948), and they trace back the true relationship between the individual if the number of markers are dense enough.

The most common genomic relationship matrix ($\mathbf{G_V}$) was proposed by VanRaden (2008). Let **M** be the marker-genotype matrix with number of individuals (**n**) and number of loci (**m**) as dimensions. The elements in the matrix **M** were coded as -1, 0 and 1 for homozygote, heterozygote and the other homozygote. The matrix **P** contains allele frequencies expressed as difference from 0.5 and multiplied by 2, then the column i of **P** was $2(p_i\text{-}0.5)$. The matrix **P** was subtracted from **M** to give **Z = M - P**. The matrix $\mathbf{G_V}$ was estimated as follow: sses: $115 due to milk yield losses, $14 because of increased mortality, and $50 for treatment-associated costs (Bar *et al*., 2008).

The variance component estimations are based on the mixed model methodologies, in which fixed and random effects are estimated simultaneously. The calculation of the matrix numerator relationship matrix (**A**) based on the complete pedigree makes possible the estimation of the additive genetic variances and the accurate prediction of the breeding values for all the animals. Recently, some studies dissected the genetic variances of complex traits by using the G matrix, also when genealogical information were not known (Hong Lee *et al.,* 2010).

Udder health: clinical mastitis.

Mastitis can be classified in clinical and subclinical cases, taking into account the presence of evident sign of inflammation and alteration in milk like clots, flakes or discolored secretion (Oliver and Calvinho, 1995; Bradley 2002). The subclinical mastitis is the main form in modern dairy herds. In addition, the mastitis etiology makes feasible another classification: infectious or contagious and non-infectious or environmental (Blowey and Edmondson, 1995). Mastitis with non-infectious etiology could be caused by mechanical or thermal traumas or chemical insults (Zhao and Lacasse 2008). About 20-35% of mastitis cases have an unknown etiology (Wellenberg *et al*., 2002).

Mastitis generally produce a decrease in milk production with different outcomings based on the pathogen causing the infection (Gröhn et al., 2004) and the genetic correlations between somatic cell score and milk yields and its components are generally around the range of -0.17 to 0.47 (i.e. Schutz et al., 1990a; Weller et al., 1992; Carlen et al., 2004).

Somatic cell count/ Somatic cell score in milk as indirect measure of mastitis resistance.

The somatic cell count in a healthy cow declines from calving to the nadir (around 60 - 90 DIM), and increases in the second part of the lactation (Wiggans and Shook, 1987; Schutz *et al*., 1990a; De Haas *et al*., 2004). In the presence of infection, SCS presents different patterns depending on the pathogen associated (De Haas *et al*., 2002b). The alterations associated with an udder infection can be detected in a period between one to three weeks before and after the bacteriological diagnosis (De Haas *et al*., 2002b; Gröhn *et al*., 2004).

The use of SCS in genetic selection has two main advantages: i) data collection is cheaper and it can be associated to routine milk recording; ii) the heritability for SCS is larger than for the direct measure of CM and the genetic correlation between CM and SCS is large with average values of 0.71 (for a review: Mrode and Swanson, 1996).

In most of the countries in the world, genetic selection for mastitis resistance is based on somatic cell score (**SCS**) (Samorè *et al*., 2006 and 2009; Heringstad *et al*., 2008), that it is the normalized measure of somatic cell count by a logarithmic transformation (Wiggans and Shook, 1987). Only a few countries consider direct information on mastitis incidence in the genetic improvements programs (Heringstad *et al*., 2001; Odergard *et al*., 2003).

Potential use of specific pathogens information for mastitis resistances.

The largest proportion of contagious mastitis is caused by bacteria such as *Staphylococcus aureus, Streptococcus dysgalactie* and *Streptococcus agalactie*. *Enterobacteriacae* as *Escherichia coli* and *Streptococcus uberi.* They represents the most common environmental pathogens causing udder infections (Bradley 2002; Pyörälä 2002; Zhao and Lacasse 2008).

When the mastitis is caused by *E. coli,* SCS level increases roughly but after 50 days the level returns to the pre-infection values; in contrast when mastitis is caused by *Staph. aureus*, *Strep. dysgalactiae*, *Strep. uberis*, and *streptococci spp,* the increases of SCC is not as fast as with the infection by *E. coli*, but the elevated levels of SCC continue lofty

for the remainder of the lactation (De Haas *et al*., 2002b). Recently the genetic parameter estimation for mastitis due to specific pathogens and SCS has been reported by Sorensen et al. (2009). The same authors concluded that it is important to consider also the pathogen causing mastitis in genetic selection for udder health in dairy cattle.

Different costs were evaluated for mastitis caused by each group of pathogens; results showed that the cost attributed to mastitis vary according to the pathogen involved (Sorensen *et al*., 2010) and by consequence it is important to attribute the correct relative emphasis in genetic selection and therefore in the selection indexes.

Genetic selection for mastitis resistance.

Genetic selection is a long term strategy to reduce the mastitis incidence in the population, and it results in a permanent change in the genetic resistance of the dairy herd (Shook, 1989). The genetic improvement is spread over all individuals of the population and it is cumulative over generations. Heritability estimated for clinical mastitis (CM) is generally low with values below 0.05 (Heringstad *et al*., 2001; Haas *et al*., 2002a; Carlen *et al*., 2004). For this reason and for the cost associated to data collection for the mastitis incidence, selection in dairy cows is often based on the use of the number of milk SCC, logarithmically transformed into SCS (Wiggans and Shook, 1987).

Low values of heritability for the incidence of clinical mastitis associated to specific pathogens were estimated in literature using either logistic or linear models (De Haas et al., 2002a). De Haas et al. 2002a estimated values of genetic correlation varying from -0.05 to 0.79 between specific pathogens mastitis and SCS or production traits (milk, fat, and protein yield). These authors considered the average of SCS on a lactation basis for 150 and 305 days of lactation. Sorensen *et al*., 2009 estimated genetic parameters for pathogen specific incidence in mastitis using a Bayesian approach and they reported values of heritability varying from 0.035 to 0.076 and always lower than values estimated for unspecific mastitis (0.109). In the same study, different values of genetic correlations resulted depending on the specific presence of a pathogen type (0.45 to 0.77). This suggested that the presence of specific pathogens in milk should be considered as specific traits in genetic selection of udder health (Sorensen *et al*., 2009).

Literature cited

Affymetrix. 2011. GeneChip® Bovine Genome Array. [Online] Address: http://media.affymetrix.com/support/technical/datasheets/bovine_datasheet.pdf. Retrieved on August 11, 2011.

Ahmadzadeh, A., F. Frago, B. Shafii, J.C. Dalton, W.J. Price and M.A. McGuire. 2009. Effect of clinical mastitis and other diseases on reproductive performance of Holstein cows. Anim Reprod Sci. 112:273-282.

Amer, P. R., and G. Banos. 2010. Implications of avoiding overlap between training and testing data sets when evaluating genomic predictions of genetic merit. J. Dairy Sci. 93:3320-3330.

Bae J.S., H. S. Cheong, L. H. Kim, S. NamGung, T. J. Park, J. Y. Chun, J. Y. Kim, C. F. A Pasaje, J. S. Lee and H. D. Shin. 2010. Identification of copy number variations and common deletion polymorphisms in cattle. BMC Genomics, 11:232. doi:10.1186/1471-2164-11-232.

Bar, D., L.W. Tauer, G. Bennett, R.N. Gonzalez, J.A. Hertl, Y.H. Schukken, H.F. Schulte, F.L. Welcome and Y.T. Gröhn. 2008. The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. J. Dairy Sci. 91:2205-2214.

Bradley, A.J., K.A. Leach, J.E. Breen, L.E. Green and M.J. Green. 2007. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. Vet. Record. 160:253-257.

Beckmann, J.S. and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor. Appl. Genet. 67**:** 35-43. doi: 10.1007/BF00303919.

Berry, D., F. Kearney, and B. Harris. 2009. Genomic Evaluation in Ireland. In Interbull International Workshop “Genomic Information in Genetic Evaluations”. Uppsala, Sweden, January 26-29, 2009.

Blowey, R, and P.W. Edmondson 1995. Mastitis Control in dairy herds. pp.29, Ipswish, Farming Press.

Bouwman A.C., Bovenhuis H., Visker M.H., van Arendonk J.A. 2011. Genome-

wide association of milk fatty acids in Dutch dairy cattle. BMC Genet. 11:12:43.

Bradley A.J. 2002. Bovine Mastitis: An Evolving Disease. The Veterinary Journal, 164, 116-128.

Buch, L.H., M.K. Sørensen, P. Berg, L.D. Pedersen and A.C. Sørensen. 2011. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. J. Anim. Breed. Genet. Article published online: 18 JUL 2011. doi: 10.1111/j.1439-0388.2011.00947.x.

Calus, M.P.L., T.H.E. Meuwissen, A.P.W. De Roos and R.F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561.

Calus, M.P.L. 2010. Genomic breeding values prediction: Methods and procedures. Animal. 4:157-164.

Carlen, E., E. Strandberg, and A. Roth. 2004. Genetic parameters for clinical mastitis, somatic cell score and production in the first there lactations of Swedish Holstein cows J. Dairy Sci. 87:3062-3070.

Cecchinato A, Penasa M, De Marchi M, Gallo L, Bittante G, Carnier P. 2011. Genetic parameters of coagulation properties, milk yield, quality, and acidity estimated using coagulating and noncoagulating milk information in Brown Swiss and Holstein-Friesian cows. J Dairy Sci.94:4205-4213.

Daetwyler, H.D., Villanueva, B and J.A. Woolliams. 2008. Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. PLoS ONE 3(10): e3395. doi:10.1371/journal.pone.0003395

Daetwyler, H. 2009. Genome-wide evaluation of populations. PhD Thesis. Wageningen University, Wageningen, The Netherlands. Pp. 192. ISBN: 978-90-8585-528-6.

David, X, A. deVries, E. Feddersen and S. Borchersen. 2010. International genomic cooperation. EuroGenomics significantly improves reliability of genomic evaluations. Proceedings of the Interbull International Workshop March 4-5, 2010, Paris, France. Interbull bulletin 41:77-78.

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel and J. M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigrees. Genetics 182: 375–385.

de Haas, Y., H.W. Barkema, and R.F. Veerkamp, 2002a. Genetic parameters of pathogen-specific incidence of clinical mastitis in dairy cows. Anim Sci. 74:233-242

de Haas, Y., H.W. Barkema, and R.F. Veerkamp, 2002b. The effect of pathogen-specific clinical mastitis on the lactation curve of somatic cell count. J. Dairy Sci. 85:1314-1323.

de Haas, Y., R.F. Veerkamp, H.W. Barkema,Y.T. Gröhn, and Y.H. Schukken 2004. Associations between Pathogen-specific cases of clinical mastitis and somatic cell count patterns. J. Dairy Sci. 87:95-105.

de Haas Y, Windig JJ, Calus MP, Dijkstra J, de Haan M, Bannink A, Veerkamp RF. 2011. Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. J Dairy Sci. 12:6122-6134.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim. Sci, 82E313-E318.

Dekkers, J.C.M. 2007. Prediction of response to marker assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331–341.

Fadista, J., Nygaard, M., L. Holm, B. Thomsen, and C. Bendixen. 2008. A snapshot of CNVs in the pig genome. PloS One 3:e3916. doi:10.1371/journal.pone.0003916.

Fadista, J., B. Thomsen, L. Holm and C. Bendixen. 2010. Copy number variation in the bovine genome. BMC Genomics 11:284. doi: 10.1186/1471-2164-11-284.

Falconer, D.S. and Mackay, T.F.C. 1996. Introduction to quantitative genetics, 4th edn. Harlow, UK: Longman.

Fernando, R.L. 1998. Genetic evaluation and selection using genotypic, phenotypic and pedigree information. In Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, NSW, Australia, Vol. 26, pp.

329–336.

Fikse, W.F. and G. Banos. 2001. Weighting Factors of Sire Daughter Information in International Genetic Evaluations. J. Dairy Sci. 84:1759-1767. doi: 10.3168/jds.S0022-0302(01)74611-5.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edin. 52:399–433.

Fontanesi L., P. Luigi Martelli, F. Beretti, V. Riggio, S. Dall'Olio, M. Colombo, R. Casadio, V. Russo and B. Portolano. 2010. An initial comparative map of copy number variations in the goat (Capra hircus) genome BMC Genomics, 11:639. doi: 10.1186/1471-2164-11-639.

Forni, S., I. Aguilar and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol**.** 43, 1. doi:10.1186/1297-9686-43-1.

Garrick, D.J., J.T. Taylor and R.L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol**.** 41:55. doi:10.1186/1297-9686-41-55.

Gianola, D., R.L. Fernando and A. Stella, 2006. Genomic assisted prediction of genetic value with semi-parametric procedures. Genetics 173: 1761–1776.

Gianola, D. and G. de los Campos. 2008. Inferring genetic values for quantitative traits non-parametrically. Genet. Res. 90:525–540.

Gianola, D., G. de los Campos, W.G. Hill, E. Manfredi and R.L. Fernando. 2009. Additive Genetic Variability and the Bayesian Alphabet. Genetics. 183:347-363.

Goddard M.E. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica 136:245-257. doi: 10.1007/s10709-008-9308-0

Goddard, M., Hayes, B. and Meuwissen, T. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed.Genet. 128:409–421. doi: 10.1111/j.1439-0388.2011.00964.x.

Gröhn, Y.T., D.J. Wilson, R.N. González, J.A. Herti, H. Schulte, G. Bennett, and Y.H. Schukken. 2004. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. J. Dairy Sci. 87:3358-3374.

Guo, G, M.S. Lund, Y. Zhang and G. Su. 2010. Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. J. Anim. Breed. Genet. 127:423-432. doi: 10.1111/j.1439-0388.2010.00878.x.

Habier, D., R.L. Fernando and J.C.M. Dekkers. 2007. The impact of genetics relationship information on Genome-assisted breeding values. Genetics 177:2389-2397.

Habier, D., R.L. Fernando, K. Kizilkaya and D.J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 2011 12:186. doi:10.1186/1471-2105-12-186.

Hayes B.J., P.M. Visscher, and M.E. Goddard. 2009a. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. 91:47–60.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain and M.E. Goddard. 2009b. Invited review: genomic selection in dairy cattle: progress and challenges. J. Dairy Sci**.** 92:433–443.

Hayes B.J., P.J. Bowman, A.C. Chamberlain, K. Verbyla, and M.E. Goddard. 2009c. Accuracy of genomic breeding values in multi-breed populations. Genet. Sel. Evol. 41:51. doi:10.1186/1297-9686-41-51.

Hazel, L.N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476.

Henderson, C.R. 1953. Estimation of variance and covariance components. Biometrics 9:226–252. doi:10.2307/3001853.

Henderson, C. R. 1976. Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. Biometrics 32:69-83.

Henderson, C.R. 1984. Applications of linear models in animal breeding. Guelph, Ontario: University of Guelph.

Heringstag, B., G. Klemetsdal, and J. Ruane. 2001. Variance components of clinical mastitis in dairy cattle – effects of trait definition and culling. Livestock Production Science 67:265-272.

Heringstag, B., G. Klementsdal, and J. Ruane. 2001. Selection responses for clinical mastitis resistance in the Norwegian cattle populations. Acta Agric. Scand. Sect. A. Anim. Sci. 51:155-160.

Heringstag, B., E. Sehested, and T. Steine. 2008. Short Communication: Correlated Responses in Somati Cell Count from Selection Against Clinical mastitis. J. Dairy Sci. 91:4437-4439.

Hobert O. 2010. The impact of whole genome sequencing on model system genetics: Get ready for the Ride. Genetics, 184:317–319.

Hong-Lee, S., M.E. Goddard, P. Visscher and J.H.J. Van der Werf. 2010. Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. Genet. Sel. Evol. 42:22. doi:10.1186/1297-9686-42-22.

Hou, Y., G.E. Liu, D.M., Bickhart, M.F. Cardone, K. Wang, L.K. Matukumalli, M. Ventura, J. Song, P.M. VanRaden, E. Kim, T.S. Sonstegard and C.P. Van Tassell. 2011. Genomic characteristics of cattle copy number variations. BMC Genomics 12:127. doi:10.1186/1471-2164-12-127.

Hugh N.Mc., T.H.E. Meuwissen, A.R. Cromie and A.K. Sonesson#.2011. Use of female information in dairy cattle genomic breeding programs. J Dairy Sci. 94: 4109-4118.

Ishwaran, H. and J.S. Rao. 2005. Spike and slab variable selection: frequentist and Bayesian strategies. The Annals of Statistics. 33:730–773. doi 10.1214/009053604000001147.

Köning S., H. Simianer and A. William. 2009. Economic evaluation of genomic breeding programs. J Dairy Sci. 92:382-391. doi:10.3168/jds.2008-1310

Legarra, A, C. Robert-Granie´, E. Manfredi and J-M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611–618.

Legarra, A., I. Aguilar and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information**.** J. Dairy Sci**.** 92:4656-4663.

Liu G.E., Y. Hou, B. Zhu, M.F. Cardone, L. Jiang, A. Cellamare, A. Mitra, L.J. Alezander, L.L. Coutinho, M.E. Dell'Aguila, L. C. Gasbarre, G. Lacalandra, R.W. Li, L. K. Matukumalli, D. Nonneman, L. C. de A. Regitano, T.P.L. Smith, J. Song, T.S. Somstegard, C.P. Van Tassell, M. Ventura, E.E. Eichler, T.G. McDaneld and J.W. Kelle. 2010. Analysis of copy number variations among diverse cattle breeds. Genome Research, 20:693-703.

Lynch, M. and Walsh, B. 1998. Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer Associates.

Loftus, R.T., D.E. MacHugh, D.G. Bradley, P.M. Sharp, and P. Cunningham. 1994. Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences of the USA 91, 2757–2761.

Lund M.S., S.P. de Ross, A.G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43:43. doi:10.1186/1297-9686-43-43.

MacNeil, M.D., Lopez-Villalobos, N and Northcutt. 2011. A prototype national cattle evaluation for feed intake and efficiency of Angus cattle. J Anim Sci. 89:3917-3923.

Malécot, G. 1948. Les mathématiques de l'hérédité. Paris: Masson and Cie, Paris. pp 1-64.

Matukumalli, L.K., C.T. Lawley, R.D. Schanbel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O'Connell, S.S. Moore, T.P.L. Smith, T.S. Sonstegard and C.P. Van Tassell. 2009. Development and characterization of high density SNP genotyping assay in cattle. Plos One 4, e5350. doi:10.1371/journal.pone.0005350.

McKay, S.D., R.D. Schnabel, B.M. Murdoch, L.K. Matukumalli, J. Aerts, W. Coppieters, D. Crews, E. Dias Neto, C.A. Gill, C. Gao, H. Mannen, Z. Wang, C.P. Van Tassell, J.L. Williams, J.F. Taylor, and S.S. Moore. 2007. Whole genome linkage

disequilibrium maps in cattle. BMC Genetics. 8:74. doi:10.1186/1471-2156-8-74.

Meuwissen, T.H.E. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35. doi:10.1186/1297-9686-41-35.

Meuwissen, T.H.E. and M. Goddard. 2010. The use of family relationships and linkage disequilibrium to impute phase and missing genotypes in up to whole-genome sequence density genotypic data. Genetics 185:1441–1449. DOI: 10.1534/genetics.110.113936.

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Meuwissen, T.H.E, T. Luan and J.A. Woolliams. 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J. Anim. Breed. Genet. doi:10.1111/j.1439-0388.2011.00966.x.

Moran D. and E. Wall. 2011. Livestock production and greenhouse gas emissions: Defining the problem and specifying solutions. Animal Frontiers 1:19-25. doi:10.2527/af.2011-0012.

Mrode, R. and G.J.T. Swanson. 1996. Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle. Anim. Breed. Abstr. 64:847-857.

Mujibi, F.D.N., J.D. Nkrumah, O.N. Durunna, P. Stothard, J. Mah, Z. Wang, J. Basarab, G. Plastow, D. H. Crews, Jr., and S. S. Moore. 2011. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. J. Anim. Sci.

Munro, G.L.; Grieve, P.A. and Kitchen, B.J. 1984. Effects of mastitis on milk yield, milk composition processing properties and yield and quality of milk products. Austr. J. Dairy Technol. 39: 7-16.

Odergard, J., G. Klemetsdal, and B. Heringstad. 2003. Genetic improvements of mastitis resistance: Validation of somatic cell score and clinical mastitis as selection criteria. J. Dairy Sci. 86:4129-4136.

Oliver, S.P. and L.F. Calvinho. 1995. Influence of inflammation on mammary

gland metabolism and milk composition. J. Anim. Sci. 73:18-33.

Ostersen, T, O.F Christensen, M. Henryon, B. Nielsen, G. Su and P. Madsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. Genet. Sel. Evol. 43:38. doi:10.1186/1297-9686-43-38.

Park, T. and G. Casella. 2008. The Bayesian LASSO. Journal of the American Statistical Association 103: 681-686.

Pyörälä S. 2002. New Strategies to Prevent Mastitis. Reprod. Dom. Anim. 37, 211 – 216.

Quass, R. L. 1976. Computing the Diagonal Elements and Inverse of a Large Numerator Relationship Matrix. Biometrics 32:949-953.

Raftery, A. and S.M. Lewis. 1992. One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. Statistical Science. 7:493-497.

Randolph, H.E. and R.E. Erwin. 1974. Influence of mastitis on properties of milk. X. Fatty acid composition. J. Dairy Sci. 57:865-868.

Ramayo-Caldas Y., A. Castelló, R. N. Pena, E. Alves, A. Mercadé, C. A Souza, A. I. Fernández, M. Perez-Enciso and J. M. Folch. 2010. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genomics, 11:593. doi:10.1186/1471-2164-11-593.

Redon R., S. Ishikawa, K.R. Fitch, L. Feuk, G.H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A, R. Carson, W. Chen, E. K. Cho, S. Dallaire, J.L. Freeman, J. R. González, M. Gratacòs, J. Huang, D. Kalaitzopoulos, D. Komura, J.R. MacDonald, C.R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang., T. Zerjal, J. Zhang, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K.W. Jones, S.W. Scherer and M.E. Hurles. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Rolfe, K.M., W.M. Snelling, M.K. Nielsen., H.C. Freetly, C.L. Ferrell and T.G.

Jenkins. 2011. Genetic and phenotypic parameter estimates for feed intake and other traits in growing beef cattle, and opportunities for selection. J Anim Sci 89:3452-3459. doi: 10.2527/jas.2011-3961

de Roos A.P.W., C. Schrooten, R.F. Veerkamp, J.A.M. van Arendonk. 2001. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J. Dairy Sci. 94:1559–1567.

Samoré, A.B., A.F.Groen. 2006. Proposal of an udder health genetic index for the Italian Holstein Friesian based on first lactation data. Ital. J. Anim. Sci. 5: 359-370.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123:218-223.

Schenkel, F.S., M. Sargolzaei, G: Kistemaker, G.B. Jansen, P. Sullivan, B.J. Van Doormaal, P.M. VanRaden and G.R. Wiggans. 2009. Reliability of Genomic Evaluation of Holstein Cattle in Canada. In Interbull International Workshop "Genomic Information in Genetic Evaluations", Uppsala, Sweden. January 26-29, 2009.

Schutz, M.M., B. Hansen, G.R. Steuernagel, and A.L. Kuck. 1990a. Variation of milk, fat, protein and somatic cells for dairy cattle. J. Dairy Sci. 73:484-493.

Schutz, M.M., B. Hansen, G.R. Steuernagel, and A.L. Kuck. 1990b. Genetic parameters for milk, fat, protein and somatic cells in Holstein. J. Dairy Sci. 73:494-502.

Shinawi M. and S.W. Cheung. 2008. The array CGH and its clinical applications. Drug Discovery Today 13:760-70.

Shook, G.E. 1989. Selection for disease resistance. J. Dairy Sci. 72:1349-1362.

Solberg, T.R., A.K. Sonesson, J.A. Woolliams and T.H.E. Meuwissen. 2008. Genomic selection using different marker types and densities. J. Anim. Sci. 86:2447–2454.

Sonstegard T.S. and C.P. Van Tassell. 2004. Bovine genomics update: making a cow jump over the moon. Genet Res 84:3-9.

Sorensen, D. and Gianola, D. 2002. Likelihood, Bayesian and MCMC methods in quantitative genetics. New York, NY:Springer.

Sorensen, L.P., P. Madsen, T. Mark and M.S. Lund. 2009a. Genetic Parameters for Pathogen-Specific Mastitis Resistances in Danish Holstein Cattle. Animal 3:5:647-656.

Sorensen, L.P., T. Mark, P. Madsen and M.S. Lund. 2009b. Genetic Correlations between Pathogen-Specific Mastitis and Somatic Cell Count in Danish Holstein. J. Dairy Sci. 92:3457-3471.

Sorensen, L.P., P. Madsen, M.K. Sorensen and S. Ostergaard. 2010. Economic values and expected effects of selecc. Animal 3:5:647-656.

Stankiewicz P. and J.R. Lupski. 2010. Structural Variation in the Human Genome and its Role in Disease. Annual Review of Medicine. 61:437–455.

Soyeurt, H., and N. Gengler. 2008. Genetic variability of fatty acids in bovine milk. Biotechnol. Agron. Soc. Environ. 12:203-210

Tsuang D.W., S.P. Millard, B. Ely, P. Chi, K. Wang, W. H. Raskind, S. Kim, Z. Brkanac and C. Yu. 2010. The Effect of Algorithms on Copy Number Variant Detection. PLoS One. 12: e14456. doi:10.1371/journal.pone.0014456.

Vandeputte-Van Messom G. and C. Burvenich.1989. Comparison of fat and cream content in normal and mastitic milk of cows. Vet Q. 11:61-64.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor and F.S. Schenkel. 2009. Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci**.** 92:16-24.

VanRaden, P.M., J.R. O'Connell, G.R. Wiggans and K.A. Weigel. .2011. Genomic evaluations with many more genotypes. Genet, Sel. Evol. 43:10. doi:10.1186/1297-9686-43-10.

K.L. Verbyla, P.J. Bowman, B.J. Hayes and M.E. Goddard. 2010. Sensitivity of genomic selection to using different prior distributions. BMC Proceedings. 4(Suppl 1):S5. doi:10.1186/1753-6561-4-S1-S5.

Wang K., M. Li, D. Hadley, R. Liu, J. Glessner, S.F.A. Grant, H. Hakonarson and M.A Bucan. 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data Genome Research 17:1665-1674.

Wang X., S. Nahashon, T. K. Feaster, A. Bohannon-Stewart and N. Adefope. 2010. An initial map of chromosomal segmental copy number variations in the chicken. BMC Genomics 2010, 11:351. doi:10.1186/1471-2164-11-351.

Wellenberg G.J., W.H.M. Van der Poel and J.T. Van Oirschot. 2002. Viral Infections and Bovine Mastitis: a review. Vet. Miocrobiol. 88: 27-45.

Weller J.I., A. Saran, and Y. Zeliger. 1992. Genetic and environmental relationship among somatic cell account, bacterial infection, and clinical mastitis. J. Dairy Sci. 75:2532-2540.

Wiggans G.R., and G.E. Shook. 1987. A lactation measure of somatic cell count. J Anim Sci 70:2666-2672.

Wray N.R., M.E. Goddard and P.M. Visscher. 2007. Prediction of individual genetic risk to disease from genome wide association studies. Genome Res. 17:1520-1528.

Wright, S. 1922. Coefficients of Inbreeding and Relationship. The American Naturalist 56:330-338.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K, Nyholt D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Michael E Goddard M.E. & Visscher, P.M. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42:565–69. doi:10.1038/ng.608.

Zhao, X, and P. Lacasse. 2008. Mammary tissue damage during bovine mastitis: causes and control. J. Anim. Sci. 86:57-65.

Zumbach, B, H. Jorjani and J. Dürr. 2010. Brown Swiss genomic evaluation. Proceedings of the Interbull International Workshop May 31 – June 4, 2010, Riga, Latvia. Interbull bulletin 42:44-51.

Chapter 2

Objectives

The genomic approach to genetic improvement of mastitis resistance in dairy cattle is the general objective of this thesis. Different aspects were here evaluated and studied and results are exposed in chapter 3 to 6.

The resistance to mastitis was studied using a specific data set on a local Italian breed, the Valdostana Red Pied cattle, where direct information on pathogens causing infection was available as the SCC information.

The genomic approach was then studied evaluating how different strategy of selection of training and test population affect accuracy of GEBV calculation.

The objectives are herein below indicated in details by chapter.

Chapter 3

The aim of this study is the estimation of genetic parameters for traits related to udder health in the Valdostana cattle breed, i.e. SCS and the presence of specific pathogens in milk, using linear (**LM**) and linear-threshold models (**ThrM)**. Different strategies of selection based on SCS values and the presence of specific pathogens in milk are considered. The objective is to reduce mastitis costs.

Chapter 4

The aim of this study was to evaluate different prior probability values, assumed for large marker effects, in the estimation of GEBV in dairy cattle.

Chapter 5

The main objective of this study was to estimate the fraction of genetic variance not explained by the 54K Illumina SNP chip using different marker-based relationship matrices. An important additional objective was to evaluate the effect of the choice of the base population on the proportion of the genetic variance addressed by genomic relationship matrices.

Chapter 6

In this study, the aim was to explore the influence of high accurate phenotypes on genomic predictions, by censoring the phenotypes thought the reliability to define the training population for genomic selection. An important objective here evaluated it was the effect of the effect of the generational overlapping between training and test population.

Chapter 3

Genetic Aspects of the Presence of Specific Pathogens in Milk and Somatic Cell Score in the Valdostana Cattle Breed

S.I Román-Ponce, C. Maltecca, M. Vevey, A. Bagnato and A.B. Samorè.

Introduction

Mastitis is a disease of the mammary gland and it is responsible for reduced milk production and milk quality, increased involuntary culling rates and discarded milk (Ahmadzadeh *et al*., 2009). Bar *et al*. (2008) estimated the average cost of a case of clinical mastitis (**CM**) to be $179: $115 due to milk yield losses, $14 because of increased mortality, and $50 for treatment-associated costs. This amount increases to $403 for cows with high expected future net returns. Mastitis is a very common disease in dairy farms (Olde-Riekerink *et al*., 2008; Bradley *et al*., 2007). In Nordic European countries all the current economic indexes assign a significant emphasis to mastitis related traits (Steine *et al*., 2008).

The decline in milk production associated with the occurrence of mastitis is well documented in literature (see e.g. Rajala-Schultz *et al.,* 1999), nonetheless the magnitude of this effect is different depending on the pathogen causing the infection (Gröhn *et al*., 2004).

Heritability of CM incidence in dairy cattle is relatively low with estimates ranging between 0.04 and 0.18 (Lin *et al*., 1989; Heringstad *et al*., 2003a and 2003b). In most analyses linear and threshold liability sire models (Zwald *et al*., 2009) have been employed, although examples of longitudinal data analysis are available (Heringstad *et al*., 2003c), particularly through the use of Poisson models (Rodrigues-Motta *et al*., 2007, Vazquez *et al*., 2009; Vallimont *et al*., 2009). Although a large number of genetic parameters estimates are available in literature for CM, only a limited number of studies considered the presence of specific pathogens causing the infection. De Haas *et al*. (2002a) estimated heritability values for the presence of specific pathogens in milk using both logistic and linear models obtaining values ranging from 0.02 to 0.10 and genetic correlations between the presence of specific pathogens and lactation averages of SCS, ranging from 0.04 to 0.63 (de Haas *et al*., 2002b). Using a Bayesian approach, Sorensen *et al*. (2009a and 2009b) estimated heritability values for the presence of specific pathogens in milk in a range between 0.04 and 0.08, and larger values for non-specific mastitis (0.11). The same authors reported various values of genetic correlations among pathogen species (from 0.45 to 0.77) and concluded that pathogens in milk should be considered as different traits in genetic selection programs.

Finally the economic value of mastitis varies with the pathogen involved (Sorensen *et al*., 2010) and therefore, in setting up selection strategies, the availability of specific information on bacterium causing the infection could be an alternative to the use of the general trait of CM event or the indirect measure of SCS.

The Valdostana is a mountain dual purpose Italian cattle breed. Milk produced is almost entirely transformed into a typical local cheese named Fontina (Ambrosoli and Pisu, 1996). Currently the breed is not selected for mastitis resistance. Nonetheless data are regularly collected for SCS for all lactating cows and a large number of milk bacteriological tests for the presence of specific pathogens are also available.

The aim of the present paper is the estimation of genetic parameters for traits related to udder health in the Valdostana cattle breed, i.e. SCS and the presence of specific pathogens in milk, using linear (**LM**) and linear-threshold models (**ThrM)**. Different strategies of selection based on SCS values and the presence of specific pathogens in milk are considered. The objective is to reduce mastitis costs.

Material and methods

Field data

Data was provided by the Italian National Breeders Association of Valdostana Cattle and was collected from 2001 to 2008. Information on somatic cell count was collected together with milk yield (**MY**), during routine milk recordings and included 802,459 test day records of 47,412 cows. Milk bacteriological tests were performed for all cows in herds where at least one case of CM was detected by veterinary services. This produced a different data set, not necessarily simultaneous to the milk and somatic cell count recording data, with information on the eventual presence of specific pathogens in individual milk but without any clinical record of the mastitis occurrence. Bacteriological milk analyses, for a total of 34,291 cows milking from 2002 to 2008, recorded the eventual presence of the following pathogens: *Staphylococcus aureus* (**STAUR**), *Streptococcus agalactie* (**STREA**), *Staphylococcus ssp.* (**STAPH***)*, *Streptococcus ssp.* (**STREP**)*, Escherichia coli* (**ECOL**), minor pathogens gram positive (**GP**), minor pathogens gram negative (**GN**) and fungi (**FUNG**). Pathogen specific data were recorded as 1 for presence and 0 for absence of the specific pathogen.

Data Editing

One record per cow was kept for the analyses as only few records (<10%) were repeated more than once in bacteriological data. At least three observations for each level of all fixed effects in the model were required. The edited data set included 23,907 records from cow's daughters of 2,500 sires and 17,573 dams with the corresponding information on pathogen presence and SCS data in adjacent days. The pedigree file consisted on five generations of ancestors extracted from the Italian Herd Book for a total of 53,244 animals. Three subpopulations are recorded in the Valdostana Herd Book: Pezzata Rossa, Pezzata Nera and Pezzata Castana. The last two groups of animals, Pezzata Nera and Castana, are considered jointly in the official national genetic evaluation and they were therefore considered as one population in this study.

Eight specific pathogens classes were included in the analysis. These were: STAUR, STREA, STAPH, STREP, ECOL, GN, GP and FUNG. Information on test-day MY and somatic cell counts were extracted from the milk recording data set choosing the closest one to the bacteriological test. Values of somatic cell counts were logarithmic transformed into SCS (Wiggans and Shook, 1987), and used as an indirect measure of pathogen unspecific mastitis. The environmental effects considered in the models were: combination of herd-year (4,701 levels), month in milk (12 levels), month of calving (12 levels), parity (five levels: from 1 to 4 and greater than 4), and breed type (two levels). Data was analyzed twice with two different models, considering the specific pathogen presence as a continuous (**LM)** or as categorical (**ThrM**) trait. In both models, all other traits were considered as linear.

The multiple trait LM for the 10 traits (8 specific pathogens, SCS and MY) has the following form:

$$\mathbf{Y} = \boldsymbol{X}b + \boldsymbol{Z}u + e,$$

where **Y** is a matrix of phenotypic records; **X** is the matrix of incidence for fixed effects (herd-year, month in milk, month of calving, parity and breed type); **Z** is the incidence matrix for animal random effect; **b** is the vector of solutions for the fixed effects; **u** is the vector with solutions for random effects, and **e** is the residual vector. Variance structure is as follow:

$$var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}\otimes\boldsymbol{G} & 0 \\ 0 & \boldsymbol{R} \end{bmatrix},$$

where **A** is the standard numerator relationship matrix based on pedigree, and **G** and **R** are the genetic and residual (co)variance matrices of order 10 x 10, respectively.

The multiple trait animal ThrM is similar to the LM with the same traits and effects, and with $\mathbf{Y}=\begin{bmatrix} y \\ \lambda \end{bmatrix}$ and λ representing a vector of unobserved liabilities for pathogen traits from binary outcome (presences or absence) and y representing a vector of observations for the continuous traits (MY and SCS). Variance structure is similar but residual variances are forced to 1 in the ThrM.

All analyses were performed with THRGIBBS1F90 software based on Gibbs sampling for mixed models (Sorensen *et al*, 1995). For the estimation of posterior means, the software POSTGIBBSF90 (Tsuruta, and Misztal, 2006) was used. The convergence of all analyses was assessed following the procedure proposed by Raftery and Lewis (1992) for a total of chain length of 300,000 iterations after a burn-in of 150,000 iterations.

Pathogen specific information and genetic response

With the final aim of reducing the costs due to mastitis in the population, different hypotheses of selection were evaluated. Effectiveness of selection for the indicator trait SCS was compared to aggregated genotypes including SCS and different combinations of the major pathogens STAUR, STREA, and ECOL. The choice of concentrating on the few major pathogens was driven by the existing conclusions in literature (Sorensen *et al.*, 2010). This choice was further strengthened by our estimation of genetic parameters for pathogens grouped by families, i.e. STAPH, STREP, GN, GP and FUNG. The family grouping did not allow us to calculate the specific correlations to individual pathogens and therefore their inclusion in the genetic selection scenarios could bias our conclusions. Genetic progress was calculated for each scenario after five generations (approximately 30 years) of selection for unspecific mastitis (using SCS), STAUR, STREA and ECOL. In order to evaluate the economic advantage of each option, the costs associated to the mastitis caused by specific or unspecific pathogens were retrieved from literature (Sorensen *et al*., 2010) and reported in Table 4. SCS was here considered as an indirect indicator for unspecific mastitis and the costs associated to SCS were considered to correspond to the costs calculated for unspecific mastitis (Sorensen *et al*., 2010). This cost was multiplied by 0.70 according to the average of

genetic correlations estimated in literature between SCS and mastitis (Mrode and Swanson, 1996). The value of each aggregation of traits in selection indexes was calculated as the sum of the genetic progress estimated for each trait multiplied by the corresponding cost

Results and discussion

Test day milk yield production was on average 13.92±4.64 kg. The average value of SCS (2.89±2.09) was low and comparable to average values previously reported for the Valdostana breed (Battaglini *et al*., 2005) and in other Italian dual-purpose breed, as i.e. Pezzata Rossa (ANAPRI, 2007).

Bacteriological data were recorded and analyzed for all cows in herds with at least one cow showing symptoms of CM. In the complete bacteriological data set a total of 68.71% of milk samples resulted positive for the presence of at least one of the pathogens considered and in some samples more than one pathogen was detected simultaneously. In detail, the pathogens were 22.9% (STAUR), 19.5% (STAPH), 17.2% (STREP), 16.7% (STREA), 5.5% (ECOL) and 4.1% (FUNG). The relative presence of pathogens was similar in the original data set and in the edited ones used in this analysis (Table 1). These percentages do not necessarily correspond to the average situation of the population considered as the data used in this study were collected from a non-random sample of the population (only when at least one case of CM is diagnosed). Furthermore it should be noted that the dataset does not include the detail of which cow was diagnosed for clinical and subclinical mastitis. This is in contrast with most of data reported in literature. Normally the frequency of pathogens is specifically related to bacteriological analyses of milk of cows with CM (Sorensen *et al*., 2009a and 2009b) or to the relative frequency of the presence of specific pathogens in data sets of positive samples (i.e. Smith *et al*., 1985).

Heritabilities

Heritability for MY was 0.21 (Table 2 and 3) from both LM and ThrM models. The SCS had a moderate heritability of ~0.08 with a value similar to the one calculated by Colleau and le Bihan-Duval (1995) from a pooling of 39 literature estimates, and similar to other more recent literature results (Rupp and Boichard, 1999; Samoré *et al*., 2001; de Haas *et al*., 2002b).

Heritability values obtained with linear and threshold models differed for most of pathogens and they were generally higher with the ThrM. Values from LM were 0.02 for STAUR, STREA, STAPH, STREP, ECOL, GP and GN, 0.03 for FUNG. While results ranged in a larger spectrum, with values from 0.02 (for ECOL and GP) to 0.09 (for STAUR and STAPH) in the ThrM.

For the pathogen presence in milk the discussion was focused on estimates obtained from the ThrM (Table 3). Estimates of binary traits from the ThrM were higher than the corresponding genetic parameter from the LM (Table 2). The largest heritability resulted for STAPH and STAUR (0.09), while the lowest values were for GN and ECOL (0.02). In literature, de Haas *et al.* (2002b) estimated 0.05 of heritability for STAPH in Dutch milk, while estimates for ECOL were 0.05 (Sorensen *et al.,* 2009) and 0.06 (de Haas *et al.,* 2002b). The moderate values of heritabilities for STAUR (0.09), STAPH (0.09), FUNG (0.09) and STREP (0.08) agreed with results of Sorensen *et al*. (2010) and suggested that these pathogens should be directly considered in the genetic selection for mastitis reduction. Similarly the heritability estimated for the pathogen STREA (0.06) supported previous findings in literature (de Haas *et al*., 2002b; Sorensen *et al*., 2009a and 2009b) and indicated its potential role in direct genetic selection for specific pathogens.

Genetic correlations

Genetic correlations between production traits estimated with both LM and ThrM models (Table 2 and 3) confirmed the antagonistic association between MY and SCS. Genetic correlations estimated with LM versus ThrM generally resulted in the same sign, but in smaller, absolute value, in LM than with ThrM. The genetic correlations between the specific presence of pathogens and SCS obtained with ThrM showed a wide range of values, spanning from a minimum of -0.304 (90% highest posterior density, HPD: -0.461; -0.135 for GP) to a maximum of 0.583 (HPD: 0.473; -0.694 for STREA). According to previous findings in literature (Detilleux *et al*., 1996; de Haas *et al*., 2002b), these large differences in genetic correlation values with SCS, both in sign and in strength, potentially confirm the existence of different patterns in milk cell increase depending on the specific pathogens causing the udder infection. The genetic correlations of SCS with STAUR, STREA, STREP, ECOL and FUNG were positive while they were negative with STAPH, GP and GN. Correlation values were higher in

ThrM, than in LM, and in agreement with most of literature results (de Haas *et al*., 2002b; Sorensen *et al.*, 2009a and 2009b). The largest genetic correlations with SCS was found for pathogens considered to be the most economically important in mastitis selection (Sorensen *et al*., 2010), i.e. STREA (0.584, HPD: 0.473; 0.684), ECOL (0.399, HPD: 0.194; 0.586), and STREP (0.346, HPD: 0.136; 0.573). These positive correlation values should indicate therefore that cows, with occurrence of infection by one of these three pathogens, should also react to the infection with a genetically determined increase in SCS. In contrast, the behavior in the presence of pathogens that are in negative correlation with SCS, as GP (-0.304, HPD: -0.461; -0.135) and STAPH (-0.206, HPD: -0.381; -0.027) should be different.

Genetic correlations among pathogen liabilities presented a wide range of values spanning from -0.723 (HPD: -0.821; -0.620) between STAUR and STAPH, to 0.522 (HPD: 0.362; 0.682) between STAUR and STREA. This suggests that the presence of different pathogens might be considered as specific traits in genetic selection and that antagonistic genetic effects might exist between pathogens. Furthermore, the infections caused by the two largest classes of bacteria causing mastitis, i.e. STAUR and STREA, are positively genetic correlated (0.522, HPD: 0.362; 0.682).

At phenotypic level, the presence of one of the two major pathogens causing mastitis, i.e. STAUR or STREA, is generally considered by veterinarians to be antagonist to the joint presence of some environmental pathogens, i.e. STAPH and STREP (Raindard and Poultrel, 1988). Nonetheless, the genetic correlations between STAUR and STREP (0.218, HPD: 0.033; 0.422) and STREA and STREP (0.253, HPD: 0.049; 0.467) were positive while negative genetic correlation were found between STAUR and STAPH (-0.723, HPD: -0.821; -0.620) and STREA and STAPH (-0.640, HPD:-0.762; -0.524). Differences between phenotypic and genetic behaviors between these pathogens are probably related to chemical and biochemical changes in the environmental habitats caused by the presence of each bacterium. These changes should, by consequence, contrast the eventual phenotypic infection by the other pathogens with an antagonism that it is only due to the environmental situation and not to the genetic background. At phenotypic levels the detection of two different strains of the major pathogens is rare (Kardarmideen and Pryce, 2009) and it is believed that different strategies are adopted by microorganisms to reduce bacterium competition in the same environment, as i.e. with the production of bacteriocins (Riley and Gordon, 1999).

The presence of ECOL was not clearly related with others pathogens, and also the positive genetic correlation with STAUR and the negative genetic correlation with STREP were not substantially different from zero. Instead, other couples of pathogens were positively related to each other, as i.e. STREA with FUNG (0.395, HPD: 0.135; 0.626), GN with STREA (0.309, HPD: 0.097; 0.482) and GP with STREA (0.262, HPD 0.093; - 0.434). Some pathogens presented negative genetic correlations, i.e. GP with STREP (-0.369, HPD:-0.459; -0.137) and STAUR and FUNG (-0.277, HPD: -0.488; -0.005). The other correlations were not different from zero probably due to the data set sampled rather than to true null correlations.

Genetic correlations among pathogens were larger in size when estimated with the ThrM rather than with the LM. Generally, the sign was the same with the two models, although some differences existed, i.e. between STREA and GP that was negative and moderate (-0.231, HPD: -0.418; -0.055]) with ThrM, and positive with the LM (0.262, HPD: 0.093; 0.434). Finally, in both models STREA presented a similar positive genetic relationship with GN (0.336 [HPD: 0.007; 0.603] in LM and 0.309 [HPD: 0.097; 0.482] in ThrM), with STREP (0.314 [HPD: 0.107; 0.506] in LM and 0.218 [HPD: 0.033; 0.422] in ThrM) and with FUNG (0.222 [HPD: 0.009; 0.436] in LM and 0.295 [HPD: 0.135; 0.626] in ThrM) and in contrast negative with STAPH (-0.278 [HPD: -0.493; -0.053] in LM and -0.640 [HPD: -0.762; -0.514] in ThrM). The values of genetic correlations were variable and there is no obvious evidence of common genetic determinism of immune responses. Nonetheless, genetic correlations were positive and moderate between bacteria of the same type and negative between the major pathogens.

Some traits included a mixture of different specific similar pathogens that were grouped by families, i.e. STAPH, STREP, GN, GP and FUNG. This was done according to Sorensen *et al.* (2009a). Nonetheless it should be noted that specific immune response could be determined by the specific pathogen within each group. Evidence of different immune responses to infection caused by gram positive and gram negative bacterium were given by Bannerman *et al*. (2004) and Sorensen *et al.* (2009a). They suggested that the immune response caused by the presence of gram positive bacterium was highly variable depending on the specific type involved, and probably a similar situation could happen also for mastitis determined by gram negative bacterium.

The genetic correlation values between MY and mastitis-specific pathogen were variable. Moderate unfavorable correlation values were found between MY and STAPH (0.331, HPD: 0.082; 0.574) and with STREA (0.280, HPD: 0.016; 0.526).

Specific Pathogens and Genetic Selection

Selection for decreased SCS is commonly used in dairy cattle populations as a mean to reduce susceptibility to mastitis (Mark *et al.,* 2002). Genetic correlation values between SCS and clinical mastitis incidence have been reported around 0.70 (Carlén *et al*., 2004; Koivula *et al.*, 2005) supporting the use of SCS as indirect indicator of unspecific mastitis, i.e. infection caused by any type of pathogens. In the current study, according to the estimates in Table 4, selection for SCS should efficiently succeed in reducing the costs associated to all type of mastitis (Table 4), i.e. unspecific mastitis (indirectly measured by SCS trait) or specific mastitis caused by two of the three major pathogens, i.e. of STREA and ECOLI. The economic value of SCS was calculated based on the costs reported by Sorensen *et al*. (2010) for unspecific mastitis and multiplied by the average value of genetic correlations between SCS and clinical mastitis reported in literature. This is an approximation that aims to cover the knowledge gap of the economic value for SCS in the Valdostana breed. Moreover, this economic value attributed to SCS considers only costs associated to the reduction in mastitis incidence although it is known that SCS level also favorably influences both the milk quality (Pantoja *et al.,* 2009) and the coagulation ability of milk for cheese-making (Cassandro *et al.,* 2008). These two further economic advantages of selecting for SCS are not considered here, but should be included in further analyses as they increase the advantages of genetic selection using SCS values.

The use of milk specific pathogen information in genetic selection to reduce mastitis resulted in various economic gains (here expressed as reduction in costs) depending on the costs associated to each type of mastitis. The best combination of traits aggregated STAUR, STREA, and ECOL (Figure 1).

With the trait SCS, the best costs reduction (Figure 1) was with STAUR and STREA, while the inclusion of ECOL resulted in a slightly smaller benefit than with the reduced option. Udder infection by ECOL presence produces great economic losses and udder damages. These type of mastitis are among the most studied (Burvenich *et al*., 2003; Steeneveld *et al*., 2011). Nevertheless the genetic background here estimated, i.e. the

genetic correlation values with SCS and other specific pathogen presences, differed from those of STAUR and STREA. Probably this depends from the different pathogenesis of ECOL (Brand *et al*., 2011; Steeneveld *et al*., 2011) but this result also suggests that the inclusion of ECOL within traits of the udder health index would be of interest. Specifically its inclusion would allow the reduction also of economically impactful ECOL mastitis. Our conclusion is therefore that an udder health index including SCS and all the major pathogen traits should probably be the best choice. Although this wider index has a smaller expected economic advantage compared to the one with only the three major pathogens, it would contribute meaningfully also to unspecific mastitis selection.

According to these preliminary analyses, it can be concluded that genetic selection to reduce the pathogen presence in milk of the Valdostana breed is possible. Furthermore the availability of information on specific pathogen incidence would improve the overall genetic gain expected with the simple use of SCS data. Nevertheless it should be pointed out that the extra economic gain obtained with the use of specific pathogen information in selection could be relatively small when compared to the selection for the trait SCS, and that the increase in costs due to bacteriological tests necessary would probably not entirely be justified. Economic analyses on costs associated to each specific udder infection, with special regards to the Valdostana breed area of breeding, are therefore warranted. This analysis should better define the relative weights of each single pathogen information in an aggregated selection and, consequently evaluate the relative importance of each pathogen causing the mastitis. Nonetheless, the indirect effect on udder infection resulting from other traits under selection, i.e. production, udder type and longevity traits, could strongly influence the relative importance of each single udder health trait. According to literature results and to previous experiences in other breeds, udder type traits are probably the most important to be included in an aggregated udder health index (Rupp and Boichard, 1999; Samoré and Groen, 2006), but further traits could be of interest, as bimodality of milk release (Samoré *et al.,* 2011), milk flow traits (Gray *et al.,* 2011), or electrical conductivity (Milner *et al.,* 1996; Kamphuis *et al.,* 2008).

Conclusions

The ThrM is the best model to be used when binary traits are considered, i.e. the presence/absence of pathogens in milk. This model is better at detecting the genetic variation. The presence of specific pathogens in milk resulted in moderate heritability values that justify their inclusion in udder health genetic selection strategies. Genetic selection can be done effectively for the common mastitis indicator of SCS. The inclusion of major pathogens causing mastitis (STAUR, STREA, and ECOL) is recommended as it improves the economic gain. Economic analyses of the values of both specific and unspecific mastitis for the Valdostana breed are warranted as they could improve the udder health index definition and the expected genetic gain in mastitis selection.

References

Ambrosoli, R. and E. Pisu. 1996. Aspetti nutrizionali del formaggio Fontina. La Rivista di Scienza dell'Alimentazione 25:393–398.

Ahmadzadeh, A., F. Frago, B. Shafii, J.C. Dalton, W.J. Price and M.A. McGuire. 2009. Effect of clinical mastitis and other diseases on reproductive performance of Holstein cows. Anim Reprod Sci. 112:273-282.

ANAPRI, 2007. L'indice di selezione della Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana. Accessed August 12th 2010. http://www.anapri.eu/index.php?option=com_content&view=article&id=68&Itemid=97
.

Bannerman, D.D., M.J. Paape, J.W. Lee, X. Zhao, J.C. Hope and P. Rainard. 2004. Escherichia coli and Staphylococcus aureus Elicit Differential Innate Immune Responses following Intramammary Infection. Clin. Diagn. Lab. Immunol. 11:463-472.

Battaglini, L., A. Mimosi, V. Malfatto, C. Lussiana and M. Bianchi. 2005. Milk yield and quality of Aosta cattle breeds in Alpine pasture.Ital. J. Anim. Sci. 4(suppl. 2):224-226.

Bar, D., L.W. Tauer, G. Bennett, R.N. Gonzalez, J.A. Hertl, Y.H. Schukken, H.F. Schulte, F.L. Welcome and Y.T. Gröhn. 2008. The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. J. Dairy Sci. 91:2205-2214.

Bradley, A.J., K.A. Leach, J.E. Breen, L.E. Green and M.J. Green. 2007. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. Vet. Record. 160:253-257.

Brand, B., A. Hartmann, D. Repsilber, B. Griesbeck-Zilch, O. Wellnitz , C. Kuhn, S. Ponsuksili, H.H. Meyer and M. Schwerin. 2011. Comparative expression profiling of E. coli and S. aureus inoculated primary mammary gland cells sampled from cows with different genetic predispositions for somatic cell score. Genet. Sel. Evol. 43:24.

Burvenich, C., V. Van Merris, J. Mehrzad, A. Diez-Fraile and L. Duchateau. 2003. Severity of E. coli mastitis is mainly determined by cow factors. Vet. Res. 34:521-564

Carlén, E., E. Strandberg and A. Roth. 2004. Genetic parameters for clinical mastitis, somatic cell score and production in the first there lactations of Swedish Holstein cows J. Dairy Sci. 87:3062-3070.

Cassandro, M., A. Comin, M. Ojala, R. Dal Zotto, M. De Marchi,L. Gallo,P. Carnier and G. Bittante. 2008. Genetic parameters of milk coagulation properties and their relationships with milk yield and quality traits in Italian Holstein cows. J. Dairy Sci. 91:371-376.

Colleau, J.J. and le Bihan-Duval, E. 1995.A simulation study of selection methods to improve mastitis resistance of dairy cows, J. Dairy Sci. 78:659-671.

de Haas, Y., H.W. Barkema and R.F. Veerkamp. 2002a. Effect of pathogen-specific clinical mastitis on the lactation curve of somatic cell count. J. Dairy Sci. 85:1314-1323.

de Haas, Y., H.W. Barkema and. R.F. Veerkamp. 2002b. Genetic parameters of pathogen-specific incidence of clinical mastitis in dairy cows. Anim. Sci. 74:233-242.

Detilleux, J.C. and P. Leroy. 1996. Indirect indicators of mastitis resistance. In: Interbull (Ed.). Proceedings of International workshop on genetic improvement of functional traits in cattle, Gembloux, Belgium.

Gray, K.A., F. Vacirca, A. Bagnato, A. Rossoni, A.B. Samoré and C. Maltecca. 2011. Genetic evaluations for measures of the milk flow curve in the Italian Brown Swiss population. J. Dairy Sci. 94:960-970.

Gröhn, Y.T., D.J. Wilson, R.N. González, J.A. Herti, H. Schulte, G. Bennett and Y.H. Schukken. 2004. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. J. Dairy Sci. 87:3358-3374.

Heringstad, B., G. Klementsdal and J. Ruane. 2001. Selection responses for clinical mastitis resistance in the Norwegian cattle populations.Acta Agric. Scand. Sect. A. Anim. Sci. 51:155-160.

Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal and K.A. Weigel. 2003a. Genetic change for clinical mastitis in Norwegian cattle: a threshold model analysis. J. Dairy Sci. 86:369-375.

Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal and K.A. Weigel. 2003b. Bivariate analysis of liability to clinical mastitis and to culling in first-lactation cows. J. Dairy Sci. 86:653-660.

Heringstad, B., Y.M. Chang, D. Gianola and G. Klemetsdal. 2003c. Genetic analysis of longitudinal trajectory of clinical mastitis in first-lactation Norwegian cattle. J. Dairy Sci. 86:2676-2683.

Kamphuis, C., D. Pietersma, R. Van der Tol, M. Wiedemann and H. Hogeveen. 2008. Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk. Comp. Electron. Agric. 62:169-181.

Kardarmideen, H.N. and J.E. Pryce. 2009. Genetic and economic relationship between somatic cell count and clinical mastitis and their use in selection for mastitis resistance in dairy cattle. Anim. Sci. 83:19-28

Lin, H.K., P.A. Oltenacu, L.D. Van Vleck, H.N. Erb and R.D. Smith. 1989. Heritabilities of and Genetic Correlations among 6 Health-Problems in Holstein Cows. J. Dairy Sci. 1989. 72:180-186.

Koivula, M., E.A. Mäntysaari, E. Negussie and T. Serenius, 2005. Genetic and phenotypic relationships among milk yield and somatic cell count before and after clinical mastitis. J Dairy Sci. 88:827-833.

Mark, T., F. Fikse, E. Emanuelson and J. Philipsson. 2002. International genetic evaluations of Holstein sires for milk somatic cell and clinical mastitis. J. Dairy Sci. 85:2384-2392.

Milner, P., K. L. Page, A.W. Walton and J.E. Hillerton. 1996. Detection of clinical mastitis by changes in electrical conductivity of foremilk before visible changes in milk. J. Dairy Sci.79, 83–86.

Olde-Riekerink, R.G., H.W. Barkema, D.F. Kelton and D.T. Scholl. 2008. Incidence rate of clinical mastitis on Canadian dairy farms. J. Dairy Sci.91:1366-1377.

Pantoja, J.C., D.J. Reinemann and P.L. Ruegg. 2009. Associations among milk quality indicators in raw bulk milk. J. Dairy Sci. 92:4978-4987.

Raftery, A. and S.M. Lewis. 1992. One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. Stat. Sci. 7:493-497.

Raindard, P. and B. Poultrel. 1988. Effect of naturally occurring intramammary infections by minor pathogens on new infections by major pathogens in cattle. Am. J. Vet. Res.49:327-329.

Rajala-Schultz, P.J., Y.T. Gröhn, C.E. McCulloch and C.L. Guard. 1999. Effects of clinical mastitis on milk mield in dairy cows. J. Dairy Sci. 82:1213–1220.

Riley, M.A. and D.M. Gordon. 1999. The ecological role of bacteriocins in bacterial competition. Trends. Microbiol. 7:129-133.

Rodrigues-Motta, M., D. Gianola, B. Heringstad, G.J. Rosa and Y.M. Chang. 2007. A zero-inflated poisson model for genetic analysis of the number of mastitis cases in Norwegian Red cows. J. Dairy Sci. 90:5306-5315.

Rupp, R. and D. Boichard. 1999. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. J. Dairy Sci. 82:2198–2204.

Samoré, A.B., A. Bagnato, F. Canavesi , S. Biffani and A.F. Groen. 2001. Breeding value prediction for SCC in Italian Holstein Friesian using a test-day repeatability model. Ital. J. Anim. Sci. 2:22-24.

Samoré, A.B. and A.F. Groen. 2006. Proposal of an udder health genetic index for the Italian Holstein Friesian based on first lactation data. Ital. J. Anim. Sci. 5:359-370.

Samoré, A.B., S.I. Roman-Ponce, F: Vacirca, E. Frigo, F. Canavesi, A. Bagnato amd C. Maltecca. 2011. Bimodality and the genetics of milk flow traits in the Italian Holstein Friesian breed. J. Dairy Sci. 94:4081-4089.

Smith, K.L., D.A. Todhunter and P.S. Schoenberger. 1985. Environmental mastitis: cause, prevalence, prevention. J. Dairy Sci. 68:1531-1553.

Steeneveld, W., T. van Werven, H.W. Barkema and H. Hogeveen. 2011. Cow-specific treatment of clinical mastitis: an economic approach. J Dairy Sci. 94:174-88.

Sorensen, D.A., S. Andersen, D. Gianola and I. Korsgaard. 1995. Bayesian inference in threshold using Gibbs sampling. Genet. Sel. Evol. 27:229-249.

Sorensen, L.P., P. Madsen, T. Mark and M.S. Lund. 2009a. Genetic Parameters for Pathogen-Specific Mastitis Resistances in Danish Holstein Cattle. Animal 3:647-656.

Sorensen, L.P., P. Madsen, T. Mark and M.S. Lund. 2009b. Genetic Correlations between Pathogen-Specific Mastitis and Somatic Cell Count in Danish Holstein. J. Dairy Sci. 92:3457-3471.

Sorensen, L.P., P. Madsen, M.K. Sorensen and S. Ostergaard. 2010. Economic values and expected effect of selection index for pathogen-specific mastitis under Danish conditions.J Dairy Sci. 93:358-369.

Steine, G., D. Kristofersson and A.G. Guttormsen. 2008. Economic evaluation of the breeding goal for Norwegian Red dairy cattle. J. Dairy Sci. 91:418-426.

Tsuruta, S. and I. Misztal. 2006. THRGIBBS1F90 for estimation of variance components with threshold and linear models.Proc. 8th WCGALP, Belo Horizonte, Brazil (Abstr).

Vallimont, J.E., C.D. Dechow, C.G. Sattler and J.S. Clay. 2009. Heritability estimates associated with alternative definitions of mastitis and correlations with somatic cell score and yield. J. Dairy Sci. 92:3402-3410.

Vazquez, A.I., K.A. Weigel, D. Gianola, D.M. Bates, M.A. Perez-Cabal, G.J. Rosa and Y.M. Chang. 2009. Poisson versus threshold models for genetic analysis of clinical mastitis in US Holsteins. J. Dairy Sci. 2009. 92:5239-5247.

Wiggans, G.R. and G.E. Shook. 1987. A lactation measure of somatic cell count. J Anim. Sci 70:2666-2672.

Zwald, N.R., K.A. Weigel, Y.M. Chang , R.D. Welper and J.S. Clay. 2004. Genetic selection for health traits using producer-recorded data. II. Genetic correlations, disease probabilities, and relationships with existing traits. J. Dairy Sci. 87:4295-4302.

TABLES

Table 1 - Number of milk tests resulted to be positive for the presence of *Staphylococcus aureus*, *Streptococcus agalactie*, *Staphylococcus ssp.*, *Streptococcus ssp.*, *Escherichia coli*, minor pathogens gram positive, minor pathogens gram negative, and fungi in the original data set and in the edited ones.

	Label	**Original data Set**	**Pathogen presence**	**Edited data Set**	**Pathogen presence**
Staphylococcus aureus	STAUR	17,837	22.9%	5,469	22.9%
Streptococcus agalactie	STREA	14,566	18.7%	3,989	16.7%
Staphylococcus ssp.	STAPH	13,719	17.6%	4,658	19.5%
Streptococcus ssp.	STREP	12,335	15.9%	4,108	17.2%
Escherichia coli	ECOL	3,275	4.2%	1,320	5.5%
Minor pathogens gram positive	GP	1,669	2.1%	551	2.3%
Minor pathogens gram negative	GN	1,381	1.8%	429	1.8%
Fungi	FUNG	3,131	4.0%	992	4.1%
Total	n	**77,768**		**23,907**	

Table 2 - Posterior mean of heritabilities (on diagonal), phenotypic (above), genetic (below) correlation values and 90% highest posterior density intervals (in brackets) for kg of milk yield (MY), somatic cell score (SCS) and pathogen incidence (STAUR= *Staphylococcus aureus*, STREA= *Streptococcus agalactie,* STAPH= *Staphylococcus ssp.*, STREP= *Streptococcus ssp.,* ECOL= *Escherichia coli*, GP= minor pathogens gram positive, GN= minor pathogens gram negative, and FUNG= fungi in milk) estimated with the linear model.

Table 2 (cont)

	MY	SCS	STAUR	STREA	STAPH	STREP	ECOL	GP	GN	FUNG
MY	**0.21**	-0.139	-0.008	-0.042	-0.004	0.006	-0.009	0.003	-0.008	-0.010
	[0.185; 0.228]									
SCS	0.267	**0.07**	0.110	0.196	-0.007	0.006	0.001	0.003	0.024	-0.004
	[0.059; 0.457]	[0.061; 0.089]								
STAUR	0.006	0.157	**0.02**	0.043	-0.124	-0.005	-0.001	-0.020	-0.016	-0.005
	[-0.301; 0.310]	[-0.082; 0.423]	[0.007; 0.026]							
STREA	0.189	0.515	-0.001	**0.02**	-0.096	-0.121	-0.023	-0.008	-0.016	0.007
	[-0.063; 0.424]	[0.361; 0.666]	[-0.331; 0.327]	[0.014; 0.027]						
STAPH	0.266	-0.037	-0.142	-0.278	**0.02**	-0.194	-0.055	-0.061	-0.025	-0.014
	[-0.014; 0.550]	[-0.249; 0.182]	[-0.524; 0.200]	[-0.493; -0.053]	[0.010; 0.024]					
STREP	0.456	0.295	0.314	0.027	-0.172	**0.02**	-0.077	0.014	-0.012	0.018
	[0.182; 0.731]	[0.070; 0.498]	[0.107; 0.506]	[-0.258; 0.296]	[-0.466; 0.104]	[0.015; 0.027]				
ECOL	-0.322	0.318	0.320	0.134	-0.039	0.000	**0.02**	-0.060	-0.041	0.002
	[-0.604; -0.036]	[0.132; 0.513	[0.024; 0.565]	[-0.107; 0.381]	[-0.305; 0.195]	[-0.273; 0.248]	[0.014; 0.028]			
GP	0.196	-0.360	-0.269	0.380	-0.231	-0.105	0.231	**0.03**	-0.011	0.011
	[-0.037; 0.442]	[-0.506; -0.196]	[-0.567; -0.007]	[0.196; 0.565]	[-0.418; -0.055]	[-0.291; 0.101]	[-0.015; 0.460]	[0.019; 0.032]		
GN	0.182	-0.041	-0.027	0.336	0.329	0.121	0.037	0.209	**0.02**	-0.001
	[-0.135 - 0.476]	[-0.258; 0.185]	[-0.279; 0.268]	[0.007; 0.603]	[0.075; 0.577]	[-0.075; 0.320]	[-0.253; 0.319]	[-0.012; 0.417]	[0.010; 0.027]	
FUNG	-0.280	0.229	-0.149	0.222	0.266	0.300	0.067	0.191	-0.082	**0.03**
	[-0.566 - 0.004]	[0.043; 0.394]	[-0.451; 0.184]	[0.009; 0.436]	[0.025; 0.487]	[0.000; 0.594]	[-0.195; 0.339]	[-0.004; 0.397]	[-0.364; 0.202]	[0.018; 0.035]

Table 3 - Posterior mean of heritabilities (on diagonal), phenotypic (above), genetic (below) correlation values and 90% highest posterior density intervals (in brackets) for kg of milk yield (MY), somatic cell score (SCS) and pathogen incidence (STAUR= *Staphylococcus aureus*, STREA= *Streptococcus agalactie*, STAPH= *Staphylococcus ssp.*, STREP= *Streptococcus ssp.*, ECOL= *Escherichia coli*, GP= minor pathogens gram positive, GN= minor pathogens gram negative, and FUNG= fungi in milk) estimated with the linear threshold model.

Table 3 (cont)

	MY	SCS	STAUR	STREA	STAPH	STREP	ECOL	GP	GN	FUNG
MY	**0.21**	-0.139	-0.011	-0.068	-0.006	0.010	-0.033	0.008	-0.018	-0.017
	[0.183; 0.228]									
SCS	0.278	**0.08**	0.157	0.278	-0.020	0.009	0.010	-0.011	0.017	0.002
	[0.074; 0.479]	[0.066; 0.95]								
STAUR	0.120	0.233	**0.09**	0.089	-0.126	0.023	0.013	-0.006	-0.007	-0.026
	[-0.119; 0.366]	[0.055; 0.441]	[0.064; 0.107]							
STREA	0.280	0.584	0.522	**0.06**	-0.073	-0.035	-0.012	0.003	0.005	0.016
	[0.016; 0.526]	[0.473; 0.694]	[0.362; 0.682]	[0.041; 0.070]						
STAPH	0.331	-0.206	-0.723	-0.640	**0.09**	-0.136	-0.025	-0.042	-0.018	-0.014
	[0.082; 0.574]	[-0.381; -0.027]	[-0.821; -0.620]	[-0.762; -0.514]	[0.070; 0.113]					
STREP	0.258	0.346	0.218	0.253	-0.583	**0.08**	-0.046	0.008	-0.017	0.018
	[-0.053; 0.583]	[0.136; 0.573]	[0.033; 0.422]	[0.049; 0.467]	[-0.711; -0.438]	[0.057; 0.085]				
ECOL	-1.014	0.399	0.145	-0.073	0.063	-0.156	**0.02**	-0.035	-0.017	0.001
	[-1.333; -0.639]	[0.194; 0.586]	[-0.066; 0.354]	[-0.296; 0.177]	[-0.134; 0.263]	[-0.340; 0.030]	[0.013; 0.024]			
GP	0.042	-0.304	0.010	0.262	0.088	-0.369	-0.266	**0.02**	-0.010	-0.002
	[-0.504; 0.438]	[-0.461; -0.135]	[-0.237; 0.314]	[0.093; 0.434]	[-0.139; 0.313]	[-0.630; -0.137]	[-0.459; 0.069]	[0.014; 0.030]		
GN	0.333	-0.089	-0.019	0.309	-0.010	-0.188	-0.185	0.207	**0.03**	0.008
	[-0.020; 0.731]	[-0.254; 0.090]	[-0.298; 0.226]	[0.097; 0.482]	[-0.270; 0.223]	[-0.387; 0.026]	[-0.415; 0.060]	[-0.008; 0.451]	[0.017; 0.035]	
FUNG	-0.090	0.251	-0.257	0.395	-0.124	0.051	0.094	0.063	0.096	**0.09**
	[-0.395; 0.254]	[-0.008; 0.491]	[-0.488; -0.005]	[0.135; 0.626]	[-0.350; 0.113]	[-0.295; 0.343]	[-0.161; 0.364]	[-0.234; 0.357]	[-0.128; 0.335]	[0.055; 0.121]

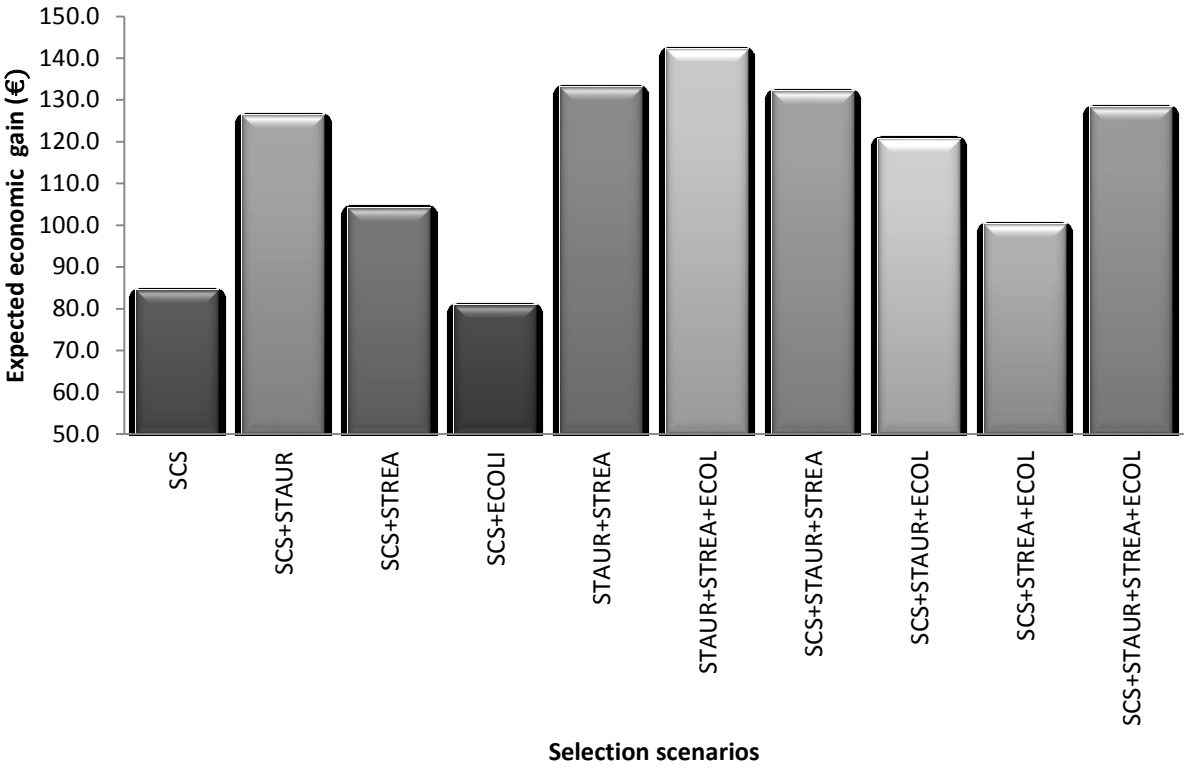
Table 4 – Various strategies of selection and the predicted genetic gain for each trait: somatic cell score (SCS) and the presence of one of the major pathogens causing mastitis in dairy cattle (STAUR= *Staphylococcus aureus*, STREA= *Streptococcus agalactie*, and ECOL= *Escherichia coli*,). Reduction in costs expected (from Sorensen *et al*., 2010) were 162€ for SCS (231€ for unspecific mastitis by 0.70 the genetic correlation between SCS and mastitis), 570€ for STAUR, 149€ for STREA, and 206€ for ECOL.

Traits in the udder health index	Genetic progress for each trait			
	SCS	STAUR	STREA	ECOL
SCS	-0.421	0.012	-0.078	-0.056
STAUR	0.027	-0.190	-0.086	0.017
STREA	-0.244	-0.122	-0.134	-0.016
ECOL	-0.367	0.050	-0.034	-0.064
SCS+STAUR	-0.382	-0.069	-0.109	-0.045
SCS+STREA	-0.411	-0.022	-0.099	-0.050
SCS+ECOL	-0.420	0.017	-0.073	-0.058
STAUR+STREA	-0.094	-0.178	-0.116	0.004
STAUR+STREA+ECOL	-0.172	-0.165	-0.122	-0.010
SCS+STAUR+STREA	-0.363	-0.084	-0.118	-0.040
SCS+STAUR+ECOL	-0.393	-0.056	-0.103	-0.049
SCS+STREA+ECOL	-0.415	-0.014	-0.094	-0.053
SCS+STAUR+STREA+ECOL	-0.377	-0.072	-0.114	-0.044

SCS= somatic cell score, STAUR= *Staphylococcus aureus*, STREA= *Staphylococcus ssp.*, ECOL= *Escherichia coli.*

FIGURE

Figure 1 Expected economic gain after 5 generations when selecting for different combinations of somatic cell score (SCS) and the major specific pathogens in milk causing mastitis, such as *Staphylococcus aureus*(STAUR), *Streptococcus agalactie*(STREA) *and, Escherichia coli* (ECOL). The economic values used were retrieved from literature.



Chapter 4

Sensitivity Analyses to Prior Probabilities on Genomic Breeding Values Prediction.

S.I. Román-Ponce, A.B. Samore', M. Dolezal, G. Banos, T.H.E. Meuwissen and A. Bagnato.

Introduction

With genomic prediction the proportion of genetic variance explained by markers on linkage disequilibrium with the quantitative trait loci (QTL) is estimated on the whole genome (Meuwissen et al, 2001). With the development of new techniques for molecular analysis, marker panels of SNP exist that may cover the whole genome (Matukumalli et al., 2009), thereby increasing the potential of genomic selection. The distribution of QTL effects has been previously described in livestock populations as a gamma distribution family member (Hayes and Godard, 2001; Xu, 2003) and genomic tools are based on it.

Genomic estimated breeding values (GEBV) are calculated basically on two different approaches. The first is based on the infinitesimal model assumption (Fisher, 1918) and literature refers to it as GBLUP. In this model, it is assumed that all SNP contribute equally to the additive genetic variances of the trait and the effects of SNP are normally distributed. Alternatively, a second approach proposes to estimate the variance estimates the variance explained by every marker. The effects of SNP are coming from an inverted chi-square distribution and the probability that a marker has a large effect is generally unknown.

Benefits of implementing genomic selection strategies in animal breeding can be associated with reduced costs (Schaeffer, 2006) increased rates of genetic gain as a consequence of the reduced generation interval and the increase of the accuracy of estimated breeding values (EBV) (Meuwissen et al., 2001). However, it is reported in literature that the values assumed for the markers effect estimation in the training population may affect the GEBV in the test population (Goddard and Hayes, 2007).

The aim of this study was to evaluate different prior probability values, assumed for large marker effects, in the estimation of GEBV in dairy cattle.

Material and methods

A total of 1089 Brown Swiss bulls were genotyped with the BovineSNP50 (Illumina). Markers on chromosome X were excluded from the analysis, leaving a total of 51,582 SNP. Editing of the markers included: the exclusion of sires with less than 90% completeness of genotyping rates (two sires deleted), deleting 583 SNP failing the test of missingness (>0.1), and deletion of 11,443 SNP with a minor allele frequency less

than 0.02. Finally, Mendelian errors were considered as missing genotypes. Editing was performed with two different software packages: SAS (SAS Inst. Inc., Cary, NC) and PLINK v1.07 (Purcell et al., 2007).The remaining data set included 1089 bulls with 39,690 SNP and a 99.35% total genotyping rate.

For the 1089 bulls considered, EBV for milk yield (MILK), fat yield (FATK), protein yield (PROTK), fat percentage (FATP), protein percentage (PROTP) and somatic cell score (SCS) were provided by the Italian Brown Cattle Breeders' Association. The training population was defined as sires born before 2001 (n=846) and the test population included all bulls born from 2001 to 2005 (n=243).

Two different models were used to estimate the markers effects (GBLUP and BayesB) on the training population. Different values of prior probabilities of SNP with large effects were tested: 39 SNP (corresponding to 0.001 of the total markers considered), 198 (0.005), 397 (0.01), 1,985 (0.05), 3,969 (0.01) and 19,845 (0.5). Results with the BayesB model were obtained after ten replicates for each probability.

The model for the markers effect estimation in the training population was as follow:

$$\mathbf{y} = \boldsymbol{\pi} + \sum_{j=1}^{m} \mathbf{X_j b_j} + \mathbf{e}$$

Where **y** is a (Tx1) vector of phenotypes with T records, **μ** is overall mean; m is total number of genotyped SNPs; $\mathbf{X_j}$ is a (Tx1) vector denoting the genotype of the individuals for markers j, $\mathbf{b_j}$ are standardized effects of the markers; and **e** is a (Tx1) vector of environmental effects.

Prediction of GEBV in the test population was done by summed the markers effect estimated as follow:

$$\text{GEBV} = \sum_{i}^{n} \mathbf{X_{ij}}\, \widehat{\mathbf{g}_j}$$

Where $\mathbf{X_i}$ is the markers genotype of the individual$_i$ for the marker j and $\widehat{\mathbf{g}_j}$ is the estimated effect of the markers j.

The comparison of markers effect solutions was done by one way analysis of variance with Tukey adjustment using the PROC GLM of SAS (SAS Inst. Inc., Cary, NC). Correlation coefficients (Pearson and Spearman) were calculated with the procedure

CORR (SAS Inst. Inc., Cary, NC) between EBV and GEBV as a measure of predictability of the phenotype and therefore as a sort of a measure of estimate of accuracy.

Results and discussion

Descriptive statistics of the BLUP-EBVs for the different traits for bulls in the training and test dataset are given in Table 1.

The rate of QTL with large effect assumed in BayesB model did not affect the value of marker effects estimated and lsmeans of QTL effects with after Tukey adjustmet did not differ among GEBV estimation methods (Table 2). Accuracies of the prediction of phenotypes (calculated as Pearson and Spearman correlation coefficients between EBV and GEBV) only slightly varied with the number of large QTL assumed with specific situation for each trait considered (Table 3). Generally values were smaller than 0.30 for PROTP, FATP and MILK and slightly bigger for FATP and PROTP (>0.40).

Spearman rank correlations (Table 4) between EBV and GEBV were slightly smaller than values in Table 3 but with similar ranges for all traits and with small differences depending on the amount of QTL with large effect assumed in the analysis.

Accuracies estimated here were generally smaller than values by Luan et al. (2009), who reported accuracies of 0.591, 0.615 and 0.617 for MILK, FATK and PROTK, respectively, when the GBLUP was used, and values of 0.577, 0.590 and 0.607, respectively, with BayesB. In contrast, with a bigger data set of 3,576 bulls in the training population and the range of accuracies were from 0.42 to 0.63, and the difference between linear and non-linear models ranged from 0.0 to 0.08, depending on the traits (VanRaden et al., 2009).

Conclusions

According to the results presented here, the value assumed for the number of SNP with a large effect did not substantially influence the estimates marker effects and the accuracies of GEBV. It seems therefore that other factors (e.g. the assumed genetic parameters and amount of phenotypic data) are more crucial at determining the GEBV accuracy achieved.

References

Goddard M.E., 2008. Genomic selection: Prediction of accuracy and maximisation of long term response, Genetica 136:245-257.

Fisher R.A., 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance, Transactions of the Royal Society of Edinburgh. 52:399-433.

Hayes B.J. and M.E. Goddard . 2001. The distribution of the effects of genes affecting quantitative traits in livestock, Genetics Selection Evolution. 33:209-22.

Hayes B.J., P.J. Bowman, A.J. Chamberlain and M.E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges, J. Dairy Sci. 92:433-443.

Luan T., J.A. Woolliams, S. Lien, M. Kent, M. Svendsen and T. Meuwissen. 2009. The accuracy of genomic selection in Norwegian Red Cattle Assessed by Cross Validation, Genetic 183:1119-1126.

Matukumalli L.K., C.T. Lawley, R.D. Schanbel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O'Connell, S.S. Moore, T.P.L. Smith, T.S. Sonstegard and C.P. Van Tassell. 2009. Development and characterization of high density SNP genotyping assay in cattle, Plos One 4:e5350.

Meuwissen, T., B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps, Genetics 157: 1819-1829.

Purcell S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly and P.C. Sham. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis, American Journal of Human Genetics, 81.

Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet, 123: 218–223.

VanRaden P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel , J.F. Taylor and F. Schenkel, 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls, J. Dairy Sci. 92:16–24.

Xu, S.Z. 2003. Estimating polygenic effects using markers of the entire genome, Genetics 163:789–801.

Table 1 Descriptive statistics (mean and Std) for estimated breeding values (EBV) and reliabilities of the Italian Brown Swiss bull population.

	Abbreviation	Scale of measure	Bulls population Total (1,089)	Training (846)	Test (243)
Fat Percentage	**FATP**	%	0.007 ± 0.165	0.16 ± 0.17	-0.02 ± 0.16
Fat Yield	**FATK**	Kg per lactation	-8.07 ± 26.12	-13.89 ± 25.68	12.22 ± 15.15
FATK Relaliability	**FATKr**	%	89.98 ± 9.63	92.42 ± 6.83	81.46 ± 12.67
Milk	**MILK**	Kg per lactation	-214.78 ± 658.83	-376.00 ± 633.26	346.33 ± 379.81
MILK Reliability	**MILKr**	%	90.53 ± 9.23	92.79 ± 6.61	82.65 ± 12.24
Protein Percentage	**PROTP**	%	-9.18E-6 ± 0.11	-0.002 ± 0.11	0.009 ± 0.10
Protein Yield	**PROTK**	Kg per lactation	-8.40 ± 23.56	-14.34 ± 22.63	12.30 ± 12.40
PROTK Relaliability	**PROTKr**	%	90.32 ± 9.36	92.64 ± 6.70	82.24 ± 12.36
Somatic Cell Score	**SCS***	Scores	57.08 ± 243.18	100.61 ± 19.14	113.36 ± 15.20
SCS Relaliability	**SCSr**	%	81.07 ± 14.14	83.25 ± 13.64	73.52 ± 13.22

SCS= (((SCS-100)/12) 0.5015) + 0.0434)

Table 2. Least-square means of marker effect solutions obtained in the Italian Brown Swiss bull population[∞].

Model	FATK	FATP	MILK	PROTK	PROTP	SCS
GBLUP	9.44E-04	2.80E-06	2.00E-02	7.30E-04	-3.05E-08	-7.69E-04
BayesB ~ 40 SNP	9.23E-04	2.95E-06	1.98E-02	7.08E-04	1.35E-07	-1.43E-03
BayesB ~ 198 SNP	9.45E-04	2.73E-06	1.87E-02	6.56E-04	2.20E-07	-3.13E-04
BayesB ~ 397 SNP	9.35E-04	2.97E-06	1.98E-02	7.24E-04	5.38E-07	-6.44E-04
BayesB ~ 1985 SNP	9.27E-04	3.01E-06	1.86E-02	7.10E-04	-1.27E-07	-9.04E-04
BayesB ~ 3969 SNP	9.64E-04	2.74E-06	1.86E-02	6.98E-04	-1.42E-07	-1.03E-03
BayesB ~ 19845 SNP	9.13E-04	3.21E-06	1.79E-02	6.60E-04	2.33E-07	-1.01E-03
Aproximate SE	1.23E-04	1.17E-06	3.17E-03	1.22E-04	7.17E-07	1.29E-03

Table 3. Pearson correlation between estimated breeding values (EBV) and genomic estimated breeding values (GEBV) in the test dataset (N=243) of the Italian Brown Swiss bull population[∞].

Model	FATK	FATP	MILK	PROTK	PROTP	SCS
GBLUP	0.255	0.410	0.188	0.146	0.557	0.462
BayesB ~ 40 SNP	0.263 ± 0.018	0.423 ± 0.028	0.165 ± 0.039	0.122 ± 0.031	0.560 ± 0.011	0.442 ± 0.043
BayesB ~ 198 SNP	0.265 ± 0.025	0.440 ± 0.018	0.182 ± 0.025	0.150 ± 0.016	0.567 ± 0.022	0.436 ± 0.011
BayesB ~ 397 SNP	0.275 ± 0.028	0.437 ± 0.021	0.194 ± 0.014	0.136 ± 0.029	0.561 ± 0.019	0.438 ± 0.015
BayesB ~ 1985 SNP	0.252 ± 0.023	0.407 ± 0.031	0.187 ± 0.019	0.148 ± 0.013	0.568 ± 0.019	0.441 ± 0.007
BayesB ~ 3969 SNP	0.243 ± 0.010	0.400 ± 0.008	0.190 ± 0.014	0.146 ± 0.017	0.541 ± 0.019	0.437 ± 0.015
BayesB ~ 19845 SNP	0.240 ± 0.024	0.380 ± 0.039	0.183 ± 0.017	0.131 ± 0.035	0.529 ± 0.022	0.434 ± 0.015

Table 4. Spearman rank correlation between estimated breeding values (EBV) and genomic estimated breeding values (GEBV) in the test data set (N=243) of the Italian Brown Swiss bull population[∞].

Model	FATK	FATP	MILK	PROTK	PROTP	SCS
GBLUP	0.271	0.410	0.163	0.162	0.564	0.425
BayesB ~ 40 SNP	0.269 ± 0.018	0.411 ± 0.028	0.143 ± 0.033	0.138 ± 0.037	0.567 ± 0.012	0.402 ± 0.042
BayesB ~ 198 SNP	0.263 ± 0.026	0.429 ± 0.018	0.165 ± 0.022	0.162 ± 0.018	0.576 ± 0.020	0.398 ± 0.012
BayesB ~ 397 SNP	0.278 ± 0.027	0.426 ± 0.018	0.174 ± 0.014	0.150 ± 0.023	0.563 ± 0.017	0.396 ± 0.015
BayesB ~ 1985 SNP	0.256 ± 0.023	0.391 ± 0.032	0.158 ± 0.018	0.157 ± 0.014	0.576 ± 0.016	0.398 ± 0.007
BayesB ~ 3969 SNP	0.254 ± 0.011	0.387 ± 0.013	0.163 ± 0.013	0.158 ± 0.017	0.546 ± 0.018	0.399 ± 0.013
BayesB ~ 19845 SNP	0.255 ± 0.023	0.373 ± 0.036	0.157 ± 0.020	0.143 ± 0.031	0.537 ± 0.021	0.394 ± 0.019

[∞]BayesB ~ 40SNP = mean of ten replicates from BayesB model with P=0.001; BayesB ~ 198SNP = mean of ten replicates from BayesB with P=0.005; BayesB ~ 397SNP = mean of ten replicates from BayesB model with P=0.01; BayesB ~ 1985SNP = mean of ten replicates from BayesB with P=0.05; BayesB ~ 3969SNP = mean of ten replicates from BayesB model with P=0.1; BayesB ~ 19845SNP = mean of ten replicates from BayesB.

Chapter 5

Estimating Missing Heritability of Complex Traits in Dairy Cattle

S.I. Román-Ponce, A.B. Samoré, M. Dolezal, A. Bagnato and T.H.E. Meuwissen.

INTRODUCTION

Due to molecular genetical advances, genome wide dense marker arrays covering all chromosomes with single nucleotide polymorphism (**SNP**) are available for livestock populations (Matukumalli *et al*., 2009). Today, several livestock populations are currently being genotyped using these arrays (Berry *et al*., 2009; Schenkel *et al*., 2009; VanRaden *et al*., 2009), with the main aim of genomic selection (**GS**) (Meuwissen *et al*., 2001). The methodology of GS predicts the genetic merit of young animals without own performance information based on marker information. Markers effects are estimated in a reference population, i.e. genotyped animals with phenotypic performances, daughter yield deviation (**DYD**) or estimated breeding values (**EBV**) derived from genetic evaluations (Calus, 2010).

The principal underlying assumption of GS is that markers are in linkage disequilibrium (**LD**) with QTL alleles (Meuwissen *et al.,* 2001; Calus *et al.,* 2008). For that reason, LD is one of the key factors that affect the accuracy of genomic breeding values (**GEBV**) (Legarra *et al.,* 2008).

Identity by descent (**IBD**) refers to alleles that descend from a common ancestor in a base population (Wright, 1922). This approach leads to the estimation of the relationship matrix based on pedigree, which is fundamental for complex traits to estimate the genetic parameters such as heritability (defined as the proportion of the phenotypic variance in a population attributable to additive genetic factors). However, the relationship matrix can also be estimated from genome-wide genetic markers i.e. panels of SNP, that can capture the additive relationship (Fernando 1998; Habier *et al.,* 2007; VanRaden, 2008; Legarra et al., 2009); which was defined as twice the coefficient of coancestry of Malécot (Malécot, 1948).

Computational methods have been developed to include genotypic data into a marker based relationship matrix. In order to estimate inbreeding and relationship coefficients the estimation of the allele frequencies in the base population is needed (VanRaden, 2008; Forni *et al.,* 2011). Recently, these relationship matrices have been used to unbiased and accurately dissect genetic variances of complex traits (Hong Lee *et al.,* 2010). The relationship matrix based on pedigree dates back to a base population, which is considered unrelated, unselected and non-inbred. The choice of the base population affects the estimate of the additive genetic variance (Van der Werf and De Boer, 1990).

The proportion of the genetic variances not addressed by markers (**C**) represent the variance that cannot be utilized by GS and the amount of C affects the maximum accuracy that can be achieved by the genomic selection. The term missing heritability (Maher, 2008) describes the fact that genome-wide association studies did not address all genetic variance estimated in complex traits (e.g. height in humans). Some potential causes of the missing heritability have been reviewed, and research strategies have been proposed to solve it, such as increase the sample size, expand the size of the studies, improve the phenotype collection, explore gene-gene interaction, change the structure of the training population (how many close relatives are included) and the use of genomic selection approaches (Manolio *et al*., 2009; Yang *et al*., 2010; Makowsky *et al*., 2011).

The main objective of this study was to estimate the fraction of genetic variance not explained by the 54K Illumina SNP chip using different marker-based relationship matrices. An important additional objective was to evaluate the effect of the choice of the base population on the proportion of the genetic variance addressed by genomic relationship matrices.

MATERIALS AND METHODS

Genomic and phenotypic data

Genotypes of Italian Brown Swiss bulls were performed with Illumina Bovine54K (Illumina Inc., San Diego, CA). All the SNPs on the X-chromosome were excluded from the analysis, and a total of 51,582 markers resulted. In the quality control, 8,892 SNP were considered as missing genotypes due to Mendelian errors. Six sires were deleted because the completeness of genotyping rates was lower than 95%. A total of 1,421 SNPs failed the test of missingness (>5%), and 14,774 SNP had a too low frequency test for minor allele frequency (<5%). Editing was performed with two different software packages: SAS (SAS Inst. Inc., Cary, NC) and PLINK v1.07 (Purcell *et al.,* 2007). After quality controls, genotypes were available for 1,086 sires with 35,706 SNP and genotypes with 99.34% total genotyping rate.

The phenotypic information consists on the EBV for fat yield (**FAT**), milk yield (**MILK**), protein yield (**PROT**), somatic cell score in milk (**SCS**), overall conformation (**OC**), stature (**STAT**), rear led side view (**RLSV**), fore udder attachment (**FUA**), rear

udder width (**RUW**), udder support (**US**), udder depth (**UD**), feet and legs (**FL**) and foot height (**FH**). These traits represent three production traits, one functional trait and ten conformation traits. Three generations of genealogical information were extracted from the official herd book (4,988 animals) and provided by the Italian Brown Cattle Breeders' Association.

Breeding values for production traits were filtered based on reliability to create four datasets as follows: animals with EBV reliability greater than 70% for each trait (**FAT70**, **MILK70**, **PROT70** and **SCS70**); animals with a minimum 90% of EBV reliability for each trait (**FAT90**, **MILK90**, **PROT90** and **SCS90**), and those animals with at least 95% of EBV reliability for each trait (**FAT95**, **MILK95**, **PROT95** and **SCS95**). The breeding values for conformation traits were filtered into two datasets as follows: those animals with more than 90% EBV reliability for each trait (**EBV90**), and those animals with at least 95% of EBV reliability for each trait (**EBV95**).

Relationship matrices: A and G

The pedigree of the genotyped sires was traced back for 3 generations and used to estimate the additive genetic relationship (**A**) with an adapted version of the procedure proposed by Meuwissen and Luo (1992) as implemented in ASREML (Gilmour *et al*., 2009).

Two genomic relationship matrices (**G**) were computed for all genotyped animals: The first genomic relationship matrix ($\mathbf{G_V}$) was computed for all the genotyped animals as proposed by VanRaden (2008). Let **M** be the marker-genotype matrix with number of individuals (**n**) and number of loci (**m**) as dimensions. The elements in the matrix **M** were coded as -1, 0 and 1 for homozygote, heterozygote and the other homozygote. The matrix **P** contains allele frequencies expressed as difference from 0.5 and multiplied by 2, then the column i of **P** was $2(p_i\text{-}0.5)$. The matrix **P** was subtracted from **M** to give **Z** = **M - P**. The matrix $\mathbf{G_V}$ was estimated as follow: $\boldsymbol{G_V} = \frac{ZZ'}{2\sum p_i(1-p_i)}$.

The second matrix was the genomic relationship matrix ($\mathbf{G_Y}$) computed as follow: $\boldsymbol{G_Y} = \frac{WW'}{m}$; where **W** is the **Z** matrix were corrected for the variance of the genotypes of the SNP as follow $W_{ij} = \frac{z_{ij}}{\sqrt{2p_j(1-p_j)}}$ (Yang *et al*., 2010).

Both the **G** and the pedigree relationship matrix, **A**, are expressed relative to a base population, i.e. an original population where all animals are assumed unrelated and non-inbred, and these populations may differ. Here the scale of the **G** was changed to that of **A**, using Wright's F-statistic (Meuwissen *et al*., 2011). We expressed the total inbreeding of animal i in the **G** matrix as: $F_{it} = G_{ii} - 1$ or $F_{it} = F_{st} + (1 - F_{st})\ F_{is}$, where F_{st} is the average inbreeding in the population, i.e. the average of the diagonal elements of **G** minus 1, and F_{is} is the inbreeding of animal i relative to the population inbreeding of F_{st}, which is calculated as: $F_{is} = \frac{(F_{it} - F_{st})}{(1 - F_{st})} = \frac{(G_{ii} - 1 - F_{st})}{(1 - F_{st})}$.

Now the population inbreeding was changed to that of **A** and the total inbreeding of individual i was calculated as: $G^*_{jj} = A_{st} + (1 - A_{st})F_{st} + 1$, where A_{st} is the average of the diagonals of **A** minus 1, G^*_{jj} is the rescaled diagonal element of **G**. Similarly, the off-diagonals were rescaled using the same F_{st} and A_{st} values. First numerator relationships were transformed to kinship, $\emptyset$, i.e. dividing the relationship by 2, and performing the base-correction on the kinship level, which is the same level as that of inbreeding, i.e. $\emptyset_{jis} = \frac{\left(\frac{G_{ji}}{2} - F_{st}\right)}{(1 - F_{st})}$, and $G^*_{ji} = 2\left[A_{st} + (1 - A_{st})\emptyset_{jis}\right]$, where $\emptyset_{jis}$ is the kinship of animal j and i relative to the population inbreeding of F_{st}.

Variance component estimation

To estimate the fraction of genetic variances captured by dense markers covering the entire genome the approach of Goddard *et al*. (2011) was used. Both **A** and **G** matrices were used to estimate the fraction of genetic variances captured by each of these matrices.

The model was fitted in ASREML-R (Butler *et al.,* 2009) as follows: $\mathbf{y} = \mu + \mathbf{Z_1}a + \mathbf{Z_2}u + e$; where, **y** is the vector of the EBV; μ is the overall mean and $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices for pedigree based and genomic random animal effects, respectively. **a** is the vector of solutions for the additive relationship matrix ($\mathbf{a} \sim N(\mathbf{0}, \mathbf{A}\sigma^2_a)$) and **u** is the vector of solutions of the genomic relationship matrix ($\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}\sigma^2_u)$). Finally, **e** is the vector of residuals ~NID ($\mathbf{0}, \mathbf{R}\sigma^2_e$).

The fraction of genetic variances not addressed by the SNP chip (**C**) was estimated as: $C = \frac{\sigma_a^2}{\sigma_g^2} = \frac{\sigma_a^2}{(\sigma_a^2+\sigma_u^2)}$; where, σ^2_g= total genetic variance, σ^2_u= variance due markers based relationship and σ^2_a= variance due to pedigree based relationship.

RESULTS

In Table 1 the descriptive statistics for FAT, MILK, PROT and SCS are showed for each trait and dataset. The average reliability (standard deviation in brackets) was ~ 90% (±9%) for FAT, PROT and MILK, and 81% (±14%) for SCS. The subsets which include sires with a minimum of 70% average reliability were similar for FAT, PROT and MILK with the average reliability of 91%. SCS had lower reliability, with values around 83%. Differences among EBV reliabilities for production and SCS disappeared on the sires subsets with at least 90 and 95% of reliability. In these datasets the average reliability was 95% (±3) for SCS and 97% (±1.5) for production traits. The descriptive statistics for each conformation trait on all datasets are shown in the Table 2. The average reliability (standard deviation in brackets), in both subsets, for all traits, was similar: 94% (±3%) and 97% (±1.5%) for EBV90 and EBV95, respectively.

For production traits, the fractions of genetic variance not explained by molecular markers (**C**) were similar across the genomic relationship matrices corrected for the base population (table 3). The estimate of C was 0.36 (FAT) and 0.34 (FAT70). The estimated C value with $\mathbf{G_Y}$ was 0.24 (±0.13) for FAT90 and the analysis with $\mathbf{G_V}$ for this trait and data set did not converge. In the subset which contains those sires with higher reliabilities (FAT95) the value of C was near or equal to zero with values of 0.05 for $\mathbf{G_V}$ and 4.22e-07 for $\mathbf{G_Y}$. The trend was similar for the other production traits, MILK and PROT (Table 3), not only for those subsets with a reliability of at least 70% (0.33±0.07), but also for the subsets with reliability greater or equal to 90%. The values of C ranged from 0.21 to 0.27 (±0.09) for MILK90 and PROT90. For MILK95 and PROT95, the amount of C values were not significant different to zero. The estimated value of C for SCS was the highest obtained in this study and ranged from 0.55 (G_V) to 0.52 (G_Y). The values of C decreased with increased data set reliability, i.e. 0.48 (±0.0) for SCS70 and 0.33 (±0.22) for SCS90. Finally, also the parameter C estimated for SCS95 was not significant different from zero.

For OC, STAT, FUA and FH, with 90% of EBV reliability, the total genetic variance was explained by $\mathbf{G_Y}$ (Table 4). The analyses did not converge for STAT95, FUA90, RUW95 and FH90 using $\mathbf{G_V}$. Although the C-values for RLSV95 (0.13±0.88) and UD95 (0.37±0.62) were substantial, they did not differ significantly from 0.

DISCUSSION

For all the traits considered here, the fraction of the genetic variance not explained by the SNPs was not significantly different from 0, when the phenotypes are >90% accurate, i.e. they approach the true genetic value. The correction of marker-based relationship and pedigree based relationship to the same population base hardly affected the amount of genetic variance explained by markers; although a small increase of genetic variance explained by the marker based matrix was observed over all the subsets (~1-2% points) for $\mathbf{G_Y}$ (results not shown). The differences in C estimates among the different genomic relationship matrices were negligible (~1% point in favor to $\mathbf{G_Y}$) in all traits and over the subsets. The only marker-based relationship matrix that converged in all analyses was $\mathbf{G_Y}$.

We estimated the fraction of the genetic variance not accounted by the Illumina 54K SNP chip (**C**) for complex dairy traits. The results showed that the estimated C values heavily depend on the accuracy of the EBV being used as y-values. Apparently, when the accuracy of the EBV increases, i.e. the correlation between EBV estimated and the true breeding value is high, the estimated fraction of the genetic variance explained by SNPs approaches the value of 100%. In contrast, when the accuracy level decreases, the family information is best explained by the **A** matrix, probably since it is calculated using the **A** matrix, which results in upwards biases of the estimates of C. The latter affects the C estimates of SCS more than those of FAT, MILK and PROT, due to the lower accuracy of SCS EBV. In the EBV90 and EBV95 data sets all the variance was captured by markers with values of C not significantly different from 0 representing the optimal situation for genomic evaluations. However in EBV70 datasets, the estimates of C were large and significantly different from 0. Since the estimates of C obtained in datasets with high accuracy were not significantly different from 0, this study found no evidence for missing heritability in dairy cattle.

If phenotypes would have been used to estimate C values instead of EBV, the upward biases mentioned above are not expected to occur. The latter is because low heritability

phenotypes result in high estimates of σ^2_e, but not inflated σ^2_a values, because family information is not being used to increase the accuracy of the phenotype (as is the case for the estimation of EBV).

The source of phenotypic information on individuals in genomic studies are often heterogeneous, i.e. they vary from individuals with reliable information, such as progeny test results, to individuals with less reliable sources e.g. individual records as in young cows. To take into account the differences in reliability of information sources it is necessary to know the value of C in order to apply the most appropriate relative weight depending on the precision of the sources of information (Garrick *et al.,* 2009).

The base population correction of the genomic relationship matrix hardly affected the proportion of genetic variances, neither the variance components estimates as in the case for pedigree based relationship matrices (Sorensen and Kennedy, 1984; Van der Werf and Boer, 1990). Moreover, this correction could make more feasible the integration of both relationship matrices **A** and **G** into a single matrix (**H**) according to Legarra *et al.* (2009), Christensen and Lund (2010) and Meuwissen *et al.* (2011).

As an alternative to estimating the total genetic variance as (s_a^2+s_u^2), i.e. the denominator of C, we also estimated the tota genetic variance using the model $\boldsymbol{y} = \mu + \mathbf{Z_1}a + e$, i.e. fitting only a pedigree based **A** matrix which is known to yield unbiased estimates of the total genetic variance. Estimates of C changed by <1% points and differences were not significant (results not reported). This alternative estimate of C is not affected by any base population correction of **G**, since if does involve an estimate of s_u^2. As mentioned above, the base population correction hardly affected the C values estimated, which is confirmed by the small changes of the estimates of C when the total genetic variance is estimated using pedigree relationships.

CONCLUSIONS

When the EBVs of the genotyped bulls are highly accurate, the fraction of the genetic variance explained by genetic markers was not significantly different from 0 for all the complex traits considered in this study. The genomic relationship matrix corrected by the heterozygosity per SNP, $\mathbf{G_Y}$ converged in all analyzes, and explained slightly more genetic variance than with the $\mathbf{G_V}$ matrix. The estimated fraction of the genetic variance explained by the Illumina 54K SNP chip was close to 100% for most traits. This

conclusion will depend of course on the SNP chip used and probably smaller SNP chips may not explain 100% of the genetic variance.

AKNOWLEDGEMENTS

The authors aknowledge the Italian Brown Cattle Breeders' Association (**ANARB**) for collecting, handling and sharing data. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 222664. ("Quantomics").

DISCLAIMER

This Publication reflects only the author's views and the European Community is not liable for any use that may be made of the information contained herein.

REFERENCES

Berry, D., F. Kearney, and B. Harris. 2009. Genomic Evaluation in Ireland. In Interbull International Workshop "Genomic Information in Genetic Evaluations". Uppsala, Sweden, January 26-29, 2009.

Butler, D.G., D.R. Cullis, A.R. Gilmour and D.J. Gogel. 2009. ASReml-R reference manual. Department of Primary industries and fisheries. Queensland. pp148.

Calus, M.P.L., T.H.E. Meuwissen, A.P.W. de Roos and R.F. Veerkamp .2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561.

Calus, M.P.L. 2010. Genomic breeding values prediction: Methods and procedures. Animal. 4:157-164.

Christensen, O.F. and M.S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. doi:10.1186/1297-9686-42-2.

Daetwyler, H. 2009. Genome-wide evaluation of populations. PhD Thesis. Wageningen University, Wageningen, The Netherlands. Pp. 192. ISBN: 978-90-8585-528-6.

Fernando, R.L. 1998. Genetic evaluation and selection using genotypic, phenotypic and pedigree information. in Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, NSW, Australia, Vol. 26, pp. 329–336.

Forni, S., I. Aguilar and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol**.** 43, 1. doi:10.1186/1297-9686-43-1.

Garrick, D.J., J.T. Taylor and R.L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol**.** 41:55. doi:10.1186/1297-9686-41-55.

Goddard, M., Hayes, B. and Meuwissen, T. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed.Genet. 128:409–421. (doi: 10.1111/j.1439-0388.2011.00964.x)

Gianola, D., G. de los Campos, W.G. Hill, E. Manfredi and R.L. Fernando. 2009. Additive Genetic Variability and the Bayesian Alphabet. Genetics. 183:347-363.

Gilmour, A.R., B.J. Gogel, B.R. Cullis and R. Thompson. 2009. ASREML User Guide Release 3.0. New South Wales Agriculture, NSW. Australia.

Habier, D, R.L. Fernando and J.C.M. Dekkers. 2007. The impact of genetics relationship information on Genome-assisted breeding values. Genetics 177:2389-2397.

Hayes, B.J., P.J. Bowman, A.J. Chamberlain and M.E. Goddard. 2009. Invited review: genomic selection in dairy cattle: progress and challenges. **J. Dairy Sci**. 92:433–443.

Hong-Lee, S., M.E. Goddard, P. Visscher and J.H.J. Van der Werf. 2010. Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. Genet. Sel. Evol. 42:22. doi:10.1186/1297-9686-42-22.

Legarra, A.C., C. Robert-Granie, E. Mandredi and J.M. Elsen. 2008. Performance of genomic Selection in Mice. Genetics. 180: 611-g18.

Legarra, A., I. Aguilar and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci**.** 92:4656-4663.

Malécot, G. 1948. Les mathématiques de l'hérédité. Paris: Masson et Cie. Vi + 1948, 63.

Matukumalli, L.K., C.T. Lawley, R.D. Schanbel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O'Connell, S.S. Moore, T.P.L. Smith, T.S. Sonstegard and C.P. Van Tassell. 2009. Development and characterization of high density SNP genotyping assay in cattle. Plos One 4, e5350. doi:10.1371/journal.pone.0005350.

Makowsky R., N.M. Pajewski, Y.C. Klimentidis, I.A. Vazquez, C.W. Duarte, D.B. Allison and G. de los Campos. 2011. Beyond Missing Heritability: Prediction of Complex Traits. PLoS Genet 7:e1002051. doi:10.1371/journal.pgen.1002051.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Golstein, L.A. Hindoff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boenhnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F.C. Mackay, S.A. McCarrol and P.M. Visscher. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753. doi:10.1038/nature08494.

Meuwissen, T.H.E. and Z. Luo .1992. Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24:305-313.

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Meuwissen, T.H.E, T. Luan and J.A. Woolliams. 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J. Anim. Breed. Genet. doi:10.1111/j.1439-0388.2011.00966.x.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller,P. Sklar, P.I.W. de Bakker, M.J. Daly and P.C. Sham. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81:559-575.

Schenkel, F.S., M. Sargolzaei, G: Kistemaker, G.B. Jansen, P. Sullivan, B.J. Van Doormaal, P.M. VanRaden and G.R. Wiggans. 2009. Reliability of Genomic Evaluation of Holstein Cattle in Canada. In Interbull International Workshop "Genomic Information in Genetic Evaluations", Uppsala, Sweden. January 26-29, 2009.

Sorensen, D.A. and B.W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci. 59:1213-1223.

Van der Werf, J.H. and I.J. de Boer. 1990. Estimation of additive genetic variance when base populations are selected. J. Anim. Sci. 68:3124-3132.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. **J. Dairy Sci**. 91:4414-4423.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor and F.S. Schenkel. 2009. Reliability of genomic predictions for North American Holstein bulls. **J. Dairy Sci**. 92:16-24.

Wright, S. 1922. Coefficients of Inbreeding and Relationship. The American Naturalist 56:330-338.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K, Nyholt D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Michael E Goddard M.E. & Visscher, P.M. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42:565–69. doi:10.1038/ng.608.

Table 1. Descriptive statistics for estimated breeding values (EBV) and reliabilities (r^2) for production traits in the genotyped sire population.

Trait	Label	Number of observation	EBV		r^2 (%)	
			Mean	Std Dev	Mean	Std Dev
Fat yield	FAT	1,086	-8.10	26.16	89.97	9.64
	FAT70	1,045	-8.93	25.95	91.45	5.66
	FAT90	741	-13.07	25.49	94.29	3.00
	FAT95	296	-16.91	27.53	97.52	1.56
Milk yield	MILK	1,086	-215.51	659.50	90.52	9.23
	MILK70	1,050	-230.82	658.14	91.84	5.46
	MILK90	772	-331.26	647.36	94.40	2.95
	MILK95	325	-459.96	667.98	97.41	1.61
Protein yield	PROT	1,086	-8.44	23.58	90.32	9.37
	PROT70	1,049	-9.08	23.44	91.67	5.57
	PROT90	759	-12.71	23.20	94.35	2.96
	PROT95	311	-16.87	24.34	97.48	1.58
Somatic cell score	SCS	1,086	0.19	0.80	81.05	14.15
	SCS70	921	0.19	0.82	85.91	7.30
	SCS90	264	0.05	0.88	94.69	3.15
	SCS95	140	0.05	0.90	97.32	1.48

Std Dev: Standard deviation.

Table 2. Descriptive statistics for estimated breeding values (EBV) and reliabilities (r^2) of the genotyped sire population.

Trait	Label	Number of observation	EBV		r^2 (%)	
			Mean	Std Dev	Mean	Std Dev
Overall conformation	OC90	214	102.49	11.51	95.35	2.83
	OC95	134	104.71	11.41	97.28	1.27
Stature	STAT90	208	100.48	11.76	95.36	2.79
	STAT95	132	101.32	12.19	97.23	1.28
Rear Leg Side View	RLSV90	127	100.06	12.09	94.77	2.60
	RLSV95	72	100.10	11.10	96.68	1.36
Fore Udder Attachment	FUA90	145	103.16	12.70	95.10	2.58
	FUA95	93	104.25	12.24	96.74	1.33
Rear Udder Width	RUW90	160	102.79	11.96	95.15	2.70
	RUW95	101	103.45	10.99	96.93	1.27
Udder Support	US90	140	102.55	13.00	94.96	2.58
	US95	86	103.35	13.05	96.69	1.32
Udder Depth	UD90	186	101.16	11.51	95.22	2.76
	UD95	118	99.16	12.03	97.05	1.29
Feet and Legs	FL90	127	101.91	9.19	94.77	2.60
	FL95	72	101.90	9.15	96.68	1.36
Foot Height	FH90	95	99.68	10.25	93.91	2.60
	FH95	40	101.38	9.29	96.55	1.15

Std Dev: Standard deviation.

Table 3. Fraction of genetic variance not addressed (**C**) ± standard error (**SE**) for production traits on dairy cattle by two relationship matrices based on dense markers corrected to the same inbreeding base population.

Label	$\mathbf{G_Y}$	$\mathbf{G_V}$
FAT	0.37 ± 0.06	0.36 ± 0.06
FAT70	0.34 ± 0.07	0.34 ± 0.07
FAT90	0.23 ± 0.10	NC
FAT95	4.22e-07 ± NA	0.05 ± 0.20
MILK	0.37 ± 0.06	0.36 ± 0.07
MILK70	0.33 ± 0.07	0.34 ± 0.07
MILK90	0.21 ± 0.09	0.27 ± 0.09
MILK95	1.96e-07 ± NA	0.07 ± 0.19
PROT	0.39 ± 0.07	0.37 ± 0.07
PROT70	0.33 ± 0.07	0.34 ± 0.07
PROT90	0.24 ± 0.09	0.33 ± 0.21
PROT95	5.66e-07 ±NA	0.24 ± 0.09
SCS	0.55 ± 0.08	0.52 ± 0.08
SCS70	0.48 ± 0.09	0.46 ± 0.09
SCS90	0.31 ± 0.22	0.33 ± 0.22
SCS95	8.55e-08 ± NA	1.05e-08 ± 1.04e-07

NA: Standard errors were not available due to parameter estimates being on the boundary of their space.

NC: Log-likelihood did not converge;$\mathbf{G_Y}$: Genomic relationship matrix as proposed by Yang *et al*., (2010); $\mathbf{G_V}$: Genomic relationship matrix as proposed by VanRaden, (2008).

Table 4. Fraction of genetic variance not addressed (**C**) ± standard error (**SE**) for conformation traits on dairy cattle by two relationship matrices based on dense markers corrected to the same inbreeding base population.

Label	$\mathbf{G_Y}$	$\mathbf{G_V}$
OC90	5.27e-07 ± NA	5.42e-07 ± NA
OC95	5.24e-07 ± NA	2.42e-07 ± NA
STAT90	4.97e-07 ± NA	0.13 ± 0.76
STAT95	7.87e-07 ± NA	NC
RLSV90	0.56 ± 0.44	0.31± 0.71
RLSV95	0.13 ± 0.88	1.18e-07 ± NA
FUA90	1.46e-07 ± NA	NC
FUA95	3.23e-06 ± NA	0.60 ± 3.06
RUW90	0.44 ± 0.37	2.71e-07 ± NA
RUW95	1.54e-06 ± NA	NC
US90	0.08 ± 0.03	0.14 ± 0.61
US95	1.90e-07 ± NA	0.36 ± 0.38
UD90	0.44 ± 0.39	0.45 ± 0.59
UD95	0.37 ± 0.62	0.29 ± 0.57
FL90	0.35 ± 0.64	7.74e-07 ± NA
FL95	6.80e-12 ± NA	3.33e-07 ±NA
FH90	5.36e-07 ± NA	NC
FH95	5.87e-05 ± NA	0.49 ± 0.37

NA: Standard errors were not available due to parameter estimates being on the boundary of their space. **NC**: Log-likelihood did not converge; $\mathbf{G_Y}$: Genomic relationship matrix as proposed by Yang *et al*., (2010); $\mathbf{G_v}$: Genomic relationship matrix as proposed by VanRaden, (2008).

Chapter 6

Effect of high accurate phenotypes in the reference population for genomic selection

S.I. Román-Ponce, A.B. Samoré, M. Dolezal, A. Bagnato and T.H.E. Meuwissen.

INTRODUCTION

The livestock populations have been genotyped using genome wide dense markers arrays covering all chromosomes with single nucleotide polymorphism (**SNP**) (Matukumalli *et al*., 2009). Genomic selection (**GS**) methodology predicts the genetic merit of young animals without own performance information based on markers information (Meuwissen *et al*., 2001). Markers effects are estimated in genotyped animals with phenotypic performances derived from genetic evaluations (Calus, 2010), named as training or reference population.

The linkage disequilibrium (**LD**) between markers and quantitative traits loci (QTL) alleles, size of the reference population, heritability of traits (defined as the proportion of the phenotypic variance in a population attributable to additive genetic factors) and number of independent chromosome segments in the population are some factors in determining the accuracy of direct genomic values (**DGV**) (Daetwyler *et al*., 2008; Calus *et al.,* 2008, Legarra *et al.,* 2008). Recently, it has been demonstrated that there is non-missing heritability in dairy cattle, in other words markers explain all the genetic variances, if high reliable enough phenotypes is used (Román-Ponce *et al*., unpublished data).

The reliabilities of genomic predictions could be evaluated by splitting the data into a set of training populations and a set of test bulls, the most common criteria has been based on birth date. An additional criteria is to select randomly bulls across birth years to integrate the test population, named cross-validation; here the effect of the degree of relatedness to the training population should be taken into account when reliabilities of genomic predictions are published (Daetwyler et al., 2009; VanRaden et al., 2009).

In this study, the aim was to explore the influence of high accurate phenotypes on genomic predictions, by censoring the phenotypes thought the reliability to define the training population for genomic selection. An important objective here evaluated it was the effect of the effect of the generational overlapping between training and test population.

MATERIALS AND METHODS

Genomic and phenotypic data

Genotypes of Italian Brown Swiss bulls derived from Illumina Bovine54K (Illumina Inc., San Diego, CA). SNP on X chromosome were not considered in the analysis. During the quality control: 9,212 SNP were considered as missing genotypes due to mendelian errors. Four sires were deleted because the completeness of genotyping rates was lower than 90%. 961 SNPs failed the test of missingness (>1%), and 10,283 SNP failed the frequency test for minor allele frequency (<0.02%). Editing was performed with two different software packages: SAS (SAS Inst. Inc., Cary, NC) and PLINK v1.07 (Purcell *et al.,* 2007). After quality controls, genotypes were available from 1,357 sires with 35,546 SNP and 99.00% total genotyping rate genotypes.

The phenotypic information was available for 1,193 sires consists on the estimated breeding values (**EBV**) for fat yield (**FAT**), milk yield (**MILK**), protein yield (**PROT**) and somatic cell score in milk (**SCS**) and it was provided by the Italian Brown Cattle Breeders' Association.

Definition of training population

Two criteria were used alone and combined to separate the genotyped sires into training and test subsets, these frames were year of birth and the reliability of the phenotypes, and were used as follow:

First definition: The population was split by the year of birth as criteria. The training population was defined as sires born before 2000 (n=935) and the test population included all bulls born after 2001 (n=258).

Second definition: Breeding values were filtered based on reliability to create two datasets as follows: those animals with more than 90% of reliability on the EBV for each trait (**EBV90**), and those animals with 95% of reliability on the EBV as lower limit for each trait (**EBV95**). The numbers of sires varying depend on the distribution of phenotypes reliabilities for each trait.

Third definition: In order to avoid generational overlapping, here the training population was defined as sires born before 2001 and these bulls were filtered based on reliability in the EBV as in the first scenario.

Estimation of DGV

GWBLUP and five replicates of BayesB models were used to estimate the markers effects on the training population for each trait as proposed by Meuwissen *et al*. (2001). The model for the markers effect estimation in the training population was as follow:

$$\mathbf{y} = \boldsymbol{\pi} + \sum_{j=1}^{m} \mathbf{X_j b_j} + \mathbf{e}$$

Where **y** is a (Tx1) vector of phenotypes with T records, **μ** is overall mean; m is total number of genotyped SNPs; $\mathbf{X_j}$ is a (Tx1) vector denoting the genotype of the individuals for markers j, $\mathbf{b_j}$ are standardized effects of the markers; and **e** is a (Tx1) vector of environmental effects.

The prediction of the DGV in both populations (training and test) was done by summed the markers effect estimated as follow:

$$\mathrm{DGV} = \sum_{i}^{n} \mathbf{X_{ij}}\, \widehat{\mathbf{g}_j}$$

Where $\mathbf{X_i}$ is the markers genotype of the individual$_i$ for the marker j and $\widehat{\mathbf{g}_j}$ is the estimated effect of the markers j.

The squared correlation coefficients of Pearson [**r**] and correlation coefficients of Spearman [**ρ]** correlation coefficients were calculated with the procedure CORR (SAS Inst. Inc., Cary, NC) between EBV and DGV as a measure of predictability of the phenotype and therefore as a sort of a measure of estimate of accuracy, in both populations (training and test).

RESULTS

Genotyped sires populations: Training and test subsets

The descriptive statistics for the whole population of genotyped sires and for the subset of sires in the training and test population are presented in Table 1. The number of bulls in the training depends on the criteria implemented to split the population. In general, fewer sires were in the training subset of sires when higher reliabilities are required in the training population, as shown in the table 1. The reliability for FAT, MILK and PROT were close to 90%, instead of 87% for SCS. These differences were not evident for those subsets which contains those animals with high accurate phenotypes (more

that 90% and 95% of reliability). The average reliability (standard deviation in brackets) in training subsets for all traits was similar, 94% (±3%) and 97% (±1.5%) for EBV90 and EBV95, respectively. FAT, MILK and PROT reliabilities in the test population for EBV90 subsets were close 80%, which were slightly higher than SCS (75%). Ten percentages point were the distances among training and test populations in the subset for EBV95

Accuracy of DGV

The predict ability of markers whether the populations was split by year of birth were the lowest in this study (Table 2), just 0.20 for SCS and for the FAT, PROT and MILK the values were ranged from 0.05 to 0.10. In the case of the definition of the training population based on reliability of the phenotypes, the squared Pearson correlation coefficients for FAT, MILK and PROT were ranged from 0.74 to 0.78 for GWBLUP, when at least 90% point of reliability for EBV were used in the training population. In the subsets which contains the highest reliabilities (>95%) the squared coefficients of correlations were ranged from 0.64 to 0.67 (Table 3). In both subsets, SCS presented the lowest accuracy 0.42 (SCS90) and 0.40 (SCS95). The predictions resulted by BayesB model varying from 0.72 to 0.77 for the subsets EBV90 and from 0.71 to 0.75 for the subset EBV95.

The squared Pearson correlation coefficient estimated between EBV and DGV, when the training population was defined first with the sires born before 2001 and then all the animals with with a minimum of 90% reliability were 5-7% points lower that the previous estimations reported in table 2. Differentness among the EBV95 subsets was close to zero (Table 4).

DISCUSSION

We estimate the DGV using two different models and three criteria to define the training population. The first criteria was split the population by year of birth, the second criterion was based on the reliability of phenotypes, and in the third, and the population was split in to step: first by year of birth, in order to avoid the generational overlapping, and as the second criterion the reliability was used to cutoff the bulls. The aim of this study was to explore the influence of high accurate phenotyping on genomic predictions with or without generational overlapping into the training population. If the

phenotype approaches the true breeding value and the generational overlapping was cancelled out; the accuracies of genomic predictions were moderately high, even if the number of animals in the training population were not the larger. For example, the accuracy for GWBLUP for FAT90 0.70 with 828 sires in the training population in contrast with 0.68 for FAT95 with 315 sires. If the generational overlapping is not nullified, the differences in the accuracies were higher (~7% points) between different cut off point of reliability. Generational overlapping did not modify the accuracy here presented for BayesB when the highest reliable phenotype (EBV95) was used, although the major number of sire in the first scenario in contrast with the second. This results lead to conclude that is feasible the use of the most reliable phenotypes to estimate the markers effects in the training population for genomic predictions, if it used to together with the year of birth to nullify the generational overlapping between the training and test population.

VanRaden et al. (2009) in North American Holstein reference bulls found higher reliability (0.23) for their total merit index (Net Merit) because of using genomic information. The largest increase in reliability was observed with high heritability traits, such as fat percentage, probably due to mutation in the DGAT1 gene (Grisart et al., 2002; Winter et al., 2002). Depending on the size of training population, other author arrive to similar conclusions with comparable reliabilities, for example New Zealand (Harris et al., 2008), the Netherlands and Flanders (De Roos et al., 2009), Australia (Hayes et al., 2009a), Ireland (Berry et al., 2009), Germany (Reinhardt et al., 2009), and Denmark (Su et al., 2010).

Currently, to increase the reliabilities of genomic predictions, the effort has been concentrated into the increment of the training population by genotyping more bulls (Wiggans et al., 2010) or international sharing of genotypes (David et al., 2010). The results achieved for total merit index in this sense with a training population with more than 16,000 Holstein sires were 0.65-0.70 (Wiggans et al., 2010; Lund et al., 2010). The results here presented comparable with these results even if the training population is quite smaller.

In this study the model comparison it was not addressed, but it is necessary to remark something some points. First at all, the accuracies of GWBLUP were higher than BayesB. The second is that the generational overlapping seems to affect GWBLUP and

not to BayesB. In both asseverations, the quantity of sires in the training population played an important role in the results, probably due to the complex assumptions in the BayesB model, and for that reason it is necessary a major number of animals in the training populations to observe the differences between both models.

CONCLUSIONS

The objective to obtain higher accuracies in the genomics predictions, direct the international effort to increase the training population by sharing genotypes of genotyping more animals such as more bulls or dams. The results of this study shown that increase of training population could be not the only one way to get up the predict ability of genomics predictions. The incorporation of the most reliable phenotypes to estimate markers effects in the training population could represent additional increments in the current accuracies on genomic evaluations.

AKNOWLEDGEMENTS

The authors' knowledge to the Italian Brown Cattle Breeders' Association (**ANARB**) for collecting, handling and sharing databases. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 222664. ("Quantomics").

DISCLAIMER

This Publication reflects only the author's views and the European Community is not liable for any use that may be made of the information contained herein.

REFERENCES

Berry, D., Kearney, F. and B. Harris. 2009. Genomic Evaluation in Ireland. In Interbull International Workshop "Genomic Information in Genetic Evaluations". Uppsala, Sweden, January 26-29, 2009.

Calus, M.P.L., T.H.E. Meuwissen, A.P.W. de Roos and R.F. Veerkamp .2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561.

Calus, M.P.L. 2010. Genomic breeding values prediction: Methods and procedures. Animal. 4:157-164.

Daetwyler H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3(10):e3395. doi:10.1371/journal.pone.0003395.

Daetwyler, H. 2009. Genome-wide evaluation of populations. PhD Thesis. Wageningen University, Wageningen, The Netherlands. Pp. 192. ISBN: 978-90-8585-528-6.

David, X., A. de Vries, E. Feddersen, and S. Borchersen, 2010. International Genomic Cooperation; EuroGenomics significantly improves reliability of genomic evaluations. Interbull Bulletin 41.

De Roos, A. P. W., C. Schrooten, E. Mullaart, S. van der Beek, G. de Jong, and W. Voskamp, 2009. Genomic selection at CRV. Interbull Bulletin 39:47-50.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell, 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222-231.

Harris, B. L., D. L. Johnson, and R. J. Spelman, 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Proc. 36th ICAR biannual

session, Niagara Falls, USA, 16-19 June, 2008, p. 325-330.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, and M. E. Goddard, 2009a. Invited review: genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433-443.

Legarra, A.C., C. Robert-Granie, E. Mandredi and J.M. Elsen. 2008. Performance of genomic Selection in Mice. Genetics. 180: 611-g18.

Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, M. Seefried, and G. Su, 2010. Improving genomic prediction by EuroGenomics collaboration. Proc. 9 of World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany, 1-6 August.

Matukumalli, L.K., C.T. Lawley, R.D. Schanbel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O'Connell, S.S. Moore, T.P.L. Smith, T.S. Sonstegard and C.P. Van Tassell. 2009. Development and characterization of high density SNP genotyping assay in cattle. Plos One 4, e5350. doi:10.1371/journal.pone.0005350.

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Reinhardt, F., Z. Liu, F. Seefried, and R. Reents, 2009. Implementation of genomic evaluation in German Holsteins. Interbull Bulletin 40:219-226.

Schenkel, F.S., M. Sargolzaei, G: Kistemaker, G.B. Jansen, P. Sullivan, B.J. Van Doormaal, P.M. VanRaden and G.R. Wiggans. 2009. Reliability of Genomic Evaluation of Holstein Cattle in Canada. In Interbull International Workshop "Genomic Information in Genetic Evaluations", Uppsala, Sweden. January 26-29, 2009.

Su, G., B. Guldbrandtsen, V. R. Gregersen, and M. Lund, 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein

population. J.Dairy Sci. 93:1175-1183.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller,P. Sklar, P.I.W. de Bakker, M.J. Daly and P.C. Sham. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81:559-575.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor and F.S. Schenkel. 2009. Reliability of genomic predictions for North American Holstein bulls. **J. Dairy Sci.** 92:16-24.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, and M. V. Silva, 2010. Increased reliability of genetic evaluations for dairy cattle in the United States from use of genomic information. Proc. 9th of World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany, 1-6 August.

Winter, A., W. Krämer, F. A. O. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G.Thaller, and R. Fries, 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol

acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. Proc. Nat. Acad. Sci. 99:9300-9305.

Table 1. Descriptive statistics for estimated breeding values (EBV) and reliabilities (r^2) of the genotyped sire population.

Trait	Label		Number of observation	EBV		r^2 (%)	
				Mean	Std Dev	Mean	Std Dev
Fat yield	**FAT**	All	1193	-10.80	28.77	89.44	10.22
	FAT90	Training	793	-13.79	25.72	94.26	2.98
	FAT90	Test	400	-4.90	33.27	79.90	12.54
	FAT95	Training	315	-17.78	27.37	97.48	1.57
	FAT90	Test	878	-8.30	28.86	86.56	10.47
Milk yield	**MILK**	All	1193	-280.88	719.17	90.01	9.83
	MILK90	Training	828	-349.79	653.20	94.36	2.93
	MILK90	Test	365	-124.54	830.07	80.13	12.50
	MILK95	Training	347	-475.84	664.04	97.36	1.61
	MILK95	Test	846	-200.91	725.97	86.99	10.20
Protein yield	**PROT**	All	1193	-10.87	25.85	89.80	9.96
	PROT90	Training	814	-13.39	23.38	94.31	2.95
	PROT90	Test	379	-5.46	29.80	80.10	12.49
	PROT95	Training	330	-17.40	24.25	97.44	1.58
	PROT95	Test	863	-8.37	26.02	86.87	10.27
Somatic cell score	**SCS**	All	1193	0.18	0.79	79.16	17.40
	SCS90	Training	281	0.03	0.86	94.63	3.18
	SCS90	Test	912	0.22	0.76	74.39	17.22
	SCS95	Training	146	0.03	0.88	97.34	1.48
	SCS95	Test	1047	0.20	0.78	76.62	17.09

Table 2. Squared coefficients of correlation of Pearson [$\mathbf{r^2}$] and Spearman[**ρ]*** between estimated breeding values (EBV) and direct genomic values (DGV) in the test population (n=258) defined by the year of birth on Italian Brown Swiss bull population.

	GWBLUP		**BAYESB**	
Trait	**r^2(EBV,DGV)**	**ρ(EBV,DGV)**	**r^2(EBV,DGV)**	**ρ(EBV,DGV)**
FAT	0.07	0.25	0.06 ± 0.004	0.26 ± 0.007
MILK	0.03	0.17	0.04 ± 0.001	0.16 ± 0.004
PROT	0.02	0.14	0.02 ± 0.003	0.14 ± 0.009
SCS	0.24	0.46	0.22 ± 0.010	0.45 ± 0.010

*Coefficients did not squared

Table 3. Squared coefficients of correlation of Pearson [$\mathbf{r^2}$] and Spearman[**ρ]*** between estimated breeding values (EBV) and direct genomic values (DGV) in the test population defined by the reliability in the phenotyping on Italian Brown Swiss bull population.

	EBV reliability > 90%					EBV reliability > 95%				
		GWBLUP		BAYESB			GWBLUP		BAYESB	
Trait	$\mathbf{n_{tr}}$	$\mathbf{r^2}$**(EBV,DGV)**	**ρ(EBV,DGV)**	$\mathbf{r^2}$**(EBV,DGV)**	**ρ(EBV,DGV)**	$\mathbf{n_{tr}}$	$\mathbf{r^2}$**(EBV,DGV)**	**ρ(EBV,DGV)**	$\mathbf{r^2}$**(EBV,DGV)**	**ρ(EBV,DGV)**
FAT	793	0.76	0.74	0.77 ± 0.010	0.73 ± 0.005	315	0.69	0.75	0.69 ± 0.003	0.75 ± 0.003
MILK	828	0.74	0.69	0.72 ± 0.004	0.68 ± 0.004	347	0.64	0.71	0.63 ± 0.003	0.71 ± 0.003
PROT	814	0.78	0.74	0.76 ± 0.002	0.73 ± 0.002	330	0.67	0.74	0.65 ± 0.005	0.74 ± 0.003
SCS	281	0.42	0.63	0.39 ± 0.006	0.60 ± 0.003	146	0.40	0.61	0.38 ± 0.003	0.60 ± 0.003

n_{tr}: number of sires in the training population; *Coefficients did not squared

Table 4. Squared coefficients of correlation of Pearson [$\mathbf{r^2}$] and Spearman[**ρ]*** between estimated breeding values (EBV) and direct genomic values (DGV) in the test population defined by year of birth <2001 and by the reliability in the phenotyping on Italian Brown Swiss bull population.

	EBV reliability > 90%					**EBV reliability > 95%**				
		GWBLUP		**BAYESB**			**GWBLUP**		**BAYESB**	
Trait	n_{tr}	**r^2(EBV,DGV)**	**ρ(EBV,DGV)**	**r^2(EBV,DGV)**	**ρ(EBV,DGV)**	n_{tr}	**r^2(EBV,DGV)**	**ρ(EBV,DGV)**	**r^2(EBV,DGV)**	**ρ(EBV,DGV)**
FAT	728	0.70	0.63	0.72 ± 0.003	0.65 ± 0.006	313	0.68	0.74	0.69 ± 0.001	0.75 ± 0.001
MILK	752	0.67	0.56	0.65 ± 0.003	0.55 ± 0.005	344	0.64	0.71	0.63 ± 0.004	0.71 ± 0.002
PROT	742	0.71	0.61	0.69 ± 0.003	0.59 ± 0.004	327	0.67	0.74	0.66 ± 0.002	0.74 ± 0.002
SCS	277	0.42	0.63	0.40 ± 0.004	0.61 ± 0.002	146	0.40	0.61	0.38 ± 0.005	0.59 ± 0.006

n_{tr}: number of sires in the training population; *Coefficients did not squared

Chapter 7

Discussion and Conclusions

The aim of this dissertation was the increase of the knowledge concerning mastitis resistance in genetic selection. Several aspects concerning mastitis resistance in genetic selection in dairy cattle were considered in order to improve the general knowledge on the topic.

In the chapter 3, binary traits reporting the presence of each specific pathogen in milk were analyzed with a threshold model to estimate genetic parameters under the classical approach of quantitative genetics. Heritabilities estimated were moderate (0.02 – 0.09) for all the specific pathogens in milk. Based on the value estimated, the author suggestion to select for mastitis resistance was the aggregation of SCS with information on the major pathogens causing mastitis, such as *Staphylococcus aureus, Streptococcus agalactie* and *Escherichia coli* in a multitrait selection.

The chapter 4 consists on a sensitivity analysis to identify the most adequate prior probability value for the number of markers with large effect to be assumed in the predictions of genomic breeding values. Despite the results suggest that the number of SNP with a large effect did not influence the estimated marker effects, the accuracies of genomic predictions were similar among all the values here studied. The recommendation in spite of the results is still to use the most appropriate number of markers with large effect resulting from previous association studies.

In the chapter 5, it was estimated the proportion of additive genetic variance addressed by dense markers in complex traits. The results when the EBVs of the genotyped bulls are highly accurate suggest that the fraction of the genetic variance explained by genetic markers is not significantly different from zero for all the traits here considered. With all the genomic relationships matrices here considered no differences resulted in the proportion of additive genetic variances explained by markers.

Finally in the chapter 6, a significant increment in the accuracy of the genomic predictions was found when the most accurate breeding values were used to estimate the SNP effects in the training population.

All the studies here presented considered genetic and genomic aspects of mastitis resistance in dairy cattle. Suggestions and experiences from the researches may contribute to the knowledge on mastitis resistance genetic selection and on genomic breeding value prediction of several traits.

The application of genomic selection in dairy cattle can enhance significantly the genetic improvement of mastitis resistance. Trait as SCC can still be used in predicting genomic resistance to mastitis, but additional genetic information may be obtained by accurate phenotyping of mastitis and pathogens causing the infection.

The genomic approach may allow an easy integration of this information into selection schemes and economic indexes thus improving efficiency of selection for mastitis resistance. In particular the possibility of addressing a specific selection for innate resistance of a specific pathogen looks a promising future possibility in livestock genetic improvement.

ACKNOWLEDGEMENTS

To the Italian National Breeders Association of Valdostana Cattle and the Italian Brown Cattle Breeders' Association for sharing the data bases in this research. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 222664. ("Quantomics").

To Prof Alessandro Bagnato from the University of Milan for the opportunity to work on this research, for his efforts and enthusiasm during these three years of hard work in Italy.

To Dra. Antonia B. Samore from the University of Milan for her hard work and enthusiasm during this first three years of collaborations.

To Prof Theo Meuwissen from the University of Life's of Sciences for all the attentions during my seven months stage in Norway.

To Giorgio Banos from the Aristotle University of Thessaloniki in Greece and to Marlies Dolezal for the advices.

To all my friends that support directly all my efforts.

To god, for let me life this experience.