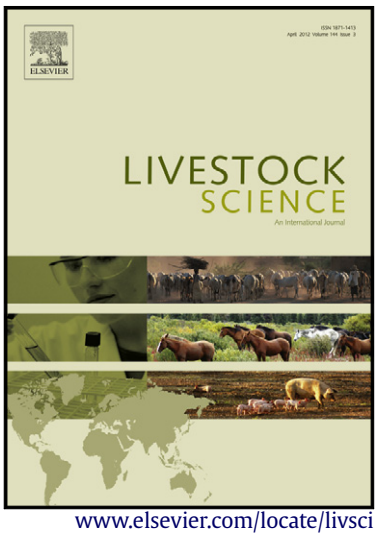

www.elsevier.com/locate/livsci

The development of genomics applied to dairy breeding

Marcos V.B. da Silva, Daniel J.A. dos Santos, Solomon A. Boison, Adam T.H. Utsunomiya, Adriana Santana do Carmo, Tad S. Sonstegard, John B. Cole, Curt P. Van Tassell

PII: S1871-1413(14)00253-4
DOI: http://dx.doi.org/10.1016/j.livsci.2014.05.017
Reference: LIVSCI2460

To appear in: *Livestock Science*

Received date: 12 May 2014
Accepted date: 15 May 2014

Cite this article as: Marcos V.B. da Silva, Daniel J.A. dos Santos, Solomon A. Boison, Adam T.H. Utsunomiya, Adriana Santana do Carmo, Tad S. Sonstegard, John B. Cole, Curt P. Van Tassell, The development of genomics applied to dairy breeding, *Livestock Science,* http://dx.doi.org/10.1016/j.livsci.2014.05.017

The development of genomics applied to dairy breeding

Marcos V.B. da Silva[a§], Daniel J. A. dos Santos[b], Solomon A. Boison[c], Adam T. H. Utsunomiya[b], Adriana Santana do Carmo[a], Tad S. Sonstegard[d], John B. Cole[d], Curt P. Van Tassell[d]

[a] Embrapa Gado de Leite - Empresa Brasileira de Pesquisa Agropecuária, Rua Eugênio do Nascimento 610 - Dom Bosco 36038-330 Juiz de Fora - MG, Brasil

[b] Universidade Estadual Paulista – UNESP, Via de Acesso Prof. Paulo Donato Castellane s/n 14884-900 Jaboticabal - SP, Brasil

[c] University of Natural Resources and Life Sciences, Gregor Mendel Str. 33, A-1180 Vienna, Austria

[d] USDA, ARS, Animal Genomics and Improvement Laboratory, 10300 Baltimore Avenue, 20705, Beltsville - MD, USA

[§]Corresponding author: e-mail: marcos.vb.silva@embrapa.br

Eugênio do Nascimento 610 - Dom Bosco 36038-330 Juiz de Fora - MG, Brasil

Tel: +55 32 3311-7459

Email addresses:

MVBS: marcos.vb.silva@embrapa.br

DJAS: daniel_jordan2008@hotmail.com

SAB: soloboan@yahoo.com

ATHU: adamtaiti@gmail.com

ASC: adriana.carmo@colaborador.embrapa.br

TSS: tad.sonstegard@ars.usda.gov

JBC: john.cole@ars.usda.gov

CPV: curt.vantassell@ars.usda.gov

Abstract

Genomic selection (GS) has profoundly changed dairy cattle breeding in the last decade and can be defined as the use of genomic breeding values (GEBV) in selection programs. The GEBV is the sum of the effects of dense DNA markers across the whole genome, capturing all the quantitative trait loci (QTL) that contribute to variation in a trait. This technology was successfully implemented in many countries as the United States, Canada, New Zealand, Australia and many other with very promising results. The GEBV reliability depends on estimation procedures and models. The different methodologies to estimate SNP effect and GEBV have been extensively tested for many research groups with very promising results. Although the GS success, many challenges still remain, including integration of GEBV into genetic evaluation programs and increasing GEBV reliability.

The aim of this review is to discuss the main aspects involved with GS, including different methodologies of imputation, SNP effect estimation, and the most important impacts of GS implementation in dairy cattle.

Keywords: dairy cattle, genomics, SNP, genomic selection

Introduction

The use of DNA marker for genetic improvement of dairy cattle was first suggested by Smith in the late 1960s (Smith, 1967), particularly for traits that are difficult to improve in conventional breeding programs because of low heritabilities or difficult-to-measure phenotypes. Affordable high-speed genotyping of large numbers of single nucleotide polymorphisms (SNP) became affordable for dairy cattle late in 2007, which permitted the development of genomic selection programs as originally described by Nejati-Javaremi et al. (1997) and expanded by Meuwissen et al. (2001).

In addition to increasing rates of genetic improvement and reducing costs of progeny testing (Meuwissen et al., 2001; Schaeffer, 2006), genomic evaluations produce estimates of the contributions of individual markers to additive genetic merit.

The rapid adoption of this technology has caused profound changes in the dairy cattle industry (Stock and Reents, 2013).

Two major technological advances were critical to the implementation and success of GS. The first was the completion of the bovine genome sequence and publication of the reference assembly, which was the basis for accelerated research progress and allowed the identification of several thousands of DNA markers, known as SNP (Elsik et al., 2009). The second one was the development of low-cost SNP chips containing thousands of markers, which enabled the estimation highly accurate breeding values when combined with phenotypic and pedigree data (Meuwissen, 2001).

In a broad sense, GS can be defined as the use of genomic breeding values (GEBV) to make selection decisions. The GEBV can be derived as the sum of the effect of markers across the genome, thereby potentially capturing all the quantitative trait loci (QTL) that contribute to variation in a trait. Reliable estimation procedures are needed for the estimation of allele substitution effects of each SNP for a trait (Hayes et al., 2009; VanRaden, 2008).

According to Schaeffer (2006), one of the main benefits of using GEBV in dairy cattle breeding programs is that selection can be made early in life, sometimes before an animal is born, reducing the generation interval. This can potentially double the rate of genetic gain. In addition, more reliable information about cows can be obtained, which may result in greater genetic progress through the dams of cows selection path (Van Tassell and Van Vleck, 2001).

The objective of this article is to review the principal aspects of GS in dairy cattle, including the most popular methods for GEBV estimation, genotype imputation, potential applications, and future perspectives.

Base of genomic analyses in Dairy Cattle: Linkage Disequilibrium, Haplotype and Persistency

In the last decade, the dairy quantitative traits began to be studied and even selected in many different breeding programs, with the aid of molecular markers. The markers can be direct, exactly marking the causative mutation of a gene, or indirect, marking regions that are nearest to the causative mutation or regions related to these mutations in only a few families (Dekkers, 2004). When working with large SNP panels to analyze quantitative traits most markers will be indirect, but in linkage with causal mutations (Dekkers, 2004). When markers are in linkage with the causative mutations, there is the possibility of recombination between the two. Recombination is a phenomenon that occurs during the formation of gametes (sperm and ovum) and involves the random exchange of genetic material between homologous chromosomes (Griffiths et al., 2009). The occurrence of recombination between two loci is proportional to the physical distance between them on the chromosome. Thus, the smaller the distance between two loci, the slower it will get to equilibrium of the expected genotype frequencies of these loci, under generations of random mating. The linkage disequilibrium (LD) measure will then indicate a nonrandom association between two loci considered, based on their genotypic and allelic frequencies (Falconer and Mackay, 1996). The main cause of LD is the "linkage" between loci because of physical proximity. Genomic selection exploits the linkage disequilibrium (LD) between markers, since it assumes that the effects of the analyzed chromosomal segments also represent the LD between the marker and a possible quantitative trait locus (QTL) (de Roos et al., 2008). The extent, distribution, and decay of LD in a population must be characterized before a genomic selection program is implemented.

Studies based on SNPs showed high LD over short distances as reported by McKay et al. (2007) and Bohamanova et al. (2011). Other authors, such as Khatkar et al. (2008) and Qanbari et al. (2010), observed $r^2 \geq 0.2$ in Holsteins for distances less than than 100 kb. Santos et al. (2013), working with a panel of 54,000 SNPs, reported r^2 of 0.15, 0.17, and 0.17 for Guzerat (n = 1025), Gyr (n = 1959), and Sindhi (n = 116), respectively. The variation in the extent of LD published depends on several factors,

including breed history and population structure (e.g., effective population size) that negatively influence LD (Hayes et al., 2003); the sample size, which can lead to overestimation in small populations (Yan et al., 2009); the density and distribution of markers; the method used to construct haplotypes; the stringency of SNP filtering (e.g., allele frequency thresholds and Hardy-Weinberg equilibrium); and the use of maternal haplotypes or both maternal and paternal haplotypes (Bohamanova et. al., 2011).

These results indicate that SNP density alone is sufficient to provide LD between chromosome segments determined for prediction of GEBV, especially when the inter-marker distances are less than 100 Kb (r^2 is moderate to high). When proposing GS in its current form, Meuwissen et al. (2001) used adjacent markers with $r^2 > 0.20$ indicating that this LD may explain the variation of the QTL. Callus et al. (2008) used simulated data to evaluate the effect of average r^2 between adjacent pairs of markers on the accuracy of genomic selection (correlation of true breeding values and to GEBV group of animals validation). They found that the accuracy of GEBV increased from 0.68 to 0.82 when the average r^2 increased from 0.1 to 0.2. Based on those results, de Roos et al. (2008) estimated that a panel of at least 50,000 SNPs would be necessary to achieve and $r^2 \geq 0.20$ between adjacent markers, which is needed to support efficient GS. Another important use for the LD is the construction of haplotypic blocks and their diversity. These blocks can be used as units for genomic analysis rather than the SNP (Calus et al. 2008), in imputation algorithms (Browning and Browning, 2009), and in genomic detection of lethal alleles (VanRanden et al., 2011). According to Khatkar et al. (2007), haplotypes are chromosomal regions of high LD and normally have low diversity, typically accounting for regions of low recombination flanked by hotpots of recombination. Generally, the structure provided by the effective size between the breeds, as well as the number of markers used, can influence the assembly of haplotypic blocks.

When LD is estimated in different populations using the same SNPs it is possible to study the persistence of phase (PS) between them. PS refers to how much

a chromosomal segment is unchanged over a given physical distance in different subpopulations, breeds, or species. This measure is based on the correlations of r^2 between two populations, along with physical distances (de Roos et al., 2008). Since PS is related to the accuracy of genome-wide association studies (GWAS) and GEBVs between populations (de Roos et al., 2008), it is possible to evaluate the feasibility of a multibreed genomic evaluation using this measure. Silva et al. (2013) working with Gyr, Guzerat, and Sindhi obtained correlations ranging from 0.40 to 0.56 for 100 kb distances, with an intense decline of PS, suggesting low efficiency for multibreed evaluation based on common SNP effects estimates for the 3 breeds. Despite the low PS observed in some cases, the increased density of the SNP panel markers, may consider high phase correlation between pairs of markers at small distances, as well as the largest LD between markers, making possible the multibreed analyses based on the same SNP effects.

Traditional marker-assisted selection and genome selection evaluation: efficiency, profitability and use of (pseudo-) phenotype

Progress in animal breeding programs is achieved through the selection of superior individuals for mating. An animal's superiority is generally based on its ranking on the basis of genetic merit. The accuracy of evaluation methods is one of the main components that determines rate of genetic gain in a population. Initially, evaluations were based only on phenotypes, i.e., the animals that had better performance were chosen for mating, or in the case of milk production, the sons of the most productive cows. Breeding values were obtained by multiplication of phenotypic deviations from the herd average with heritability.

In the latter half of the 20th century, selection index methodology was introduced by Hazel and Lush (1943). This methodology considered the relationship between phenotypic measures, as well as the genetic relationships between animals with phenotypes (selection criteria to be used - in the present left hand) and the individuals

being evaluated (objective selection - present in the right hand). With this method it was possible to combine many sources of information into a single breeding objective. In the indices the main properties would decrease the prediction error, maximizing the correlation between the estimated and true (accurate) genetic value and maximizing the probability of correct classification for the predicted genetic value. Thus, there was an increase in accuracy by aggregating information collaterally other animals. From this methodology we started to give greater importance to the pedigree of animals for use in analysis beyond parent - offspring relationships.

With the development of mixed model methods by Henderson (1949) genetic evaluations began to provide more accurate estimates of breeding value. First, through sires model that considered sire-progeny relationships, and then through the animal model, which considered all known relationships among animals in the pedigree. Using this methodology it is possible to simultaneously estimate fixed effects (BLUE - Best Linear Unbiased Estimator) and random (BLUP - Best Linear Unbiased Prediction). Thus, the BLUP solution is obtained for all animals present in the pedigree. This methodology has similar statistical properties to selection index, but directly produces estimated breeding values, unlike selection index, in which index weights and breeding values are produced in separate steps. Estimated breeding values (EBV) were widely adopted as a selection tool in breeding programs, where they are commonly presented as predicted transmitting abilities (PTA), which are one-half of EBV.

The use of molecular marker information to increase accuracy and reduce generation intervals has been studied in recent decades, and implemented in a limited fashion in some breeding programs. Marker-assisted selection (MAS) was applied in dairy cattle for the pre-selection of animals, and to select young bulls for entry intob progeny testing programs (Kashi et al., 1990; Mackinnon and Georges, 1998). MAS simultaneously uses phenotypic information and data about molecular markers in LD with QTLs, and was adopted to increase annual genetic gain for traits of economic importance in several animal species (Dekkers, 2004). In MAS, BLUP estimates of total

genetic value are obtained that include marker information as fixed or random effects (Dekkers, 2004), or through an index that combines the two sources of information using weights can be changed based on the selection objective (Dekkers and van Arendonk, 1998).

Other molecular alternatives are being widely studied. The first recognizable presentation of genomic selection was made by Nejati-Javaremi et al. (1997), and the approach was expanded and popularized by Meuwissen et al. (2001). However, there was considerable lag between the description of the concept and its widespread adoption, which did not occur until panels with thousands of single nucleotide polymorphisms (SNPs) distributed across the the bovine genome became available (Van Tassell et al. 2008). SNPs are the most abundant DNA polymorphisms in the genome, and they have become preferred over other types of molecular markers because they have low mutation rates and genotypes can easily be read automatically (Romualdi et al. 2002). In GS, the central idea is to not use specific markers for QTLs, but to use a large number of markers distributed throughout the genome. When many thousands of markers are used it can reasonably be assumed that there are always markers located near causal variants, which means that there are SNP in LD with the QTL (de Ross et al., 2008). The additive genetic merit of an animal can then be decomposed into a contribution from the markers and a polygenic component that accounts for the variation not explained by the markers. The marker and polygenic effects can be estimated using statistical models similar to those used for breeding value estimation, and performance, pedigree, and genotype information can be combined into genomic breeding values (Meuwissen et al., 2001; VanRaden, 2008). Cole et al. (2009) confirmed than an infinitesimal model is appropriate for most traits of interest in dairy production, and showed that there are few QTL in the traditional sense (loci that explain large proportions of phenotypic variance).

Genomic selection does not have the same limitations as MAS, and GS compared to BLUP provides: 1) predictions of breeding values with greater accuracy,

particularly for traits that are expressed in one sex or are of low heritability, 2) lower rates of inbreeding (lesser tendency for family selection), 3) antecipation the selection process in the case of measured characteristics later in life animals, and 4) facilitate the evaluation of the traits of difficult to measure or high cost (Daetwyler et al., 2007; Dekkers, 2007; Muir, 2007; Meuwissen, 2007).

Genomic selection has increased the rate of genetic gain in livestock (Weigel et al., 2010). The increase in the accuracy of genomic predictions is best observed in young animals, with no significant changes in the already proven bulls (Schaeffer, 2006). Proofs are used for pre-selection for progeny testing, and also for selection of animals when they are selected by GEBV in the total genomic evaluation. In didactic scheme, the genomic proofs for dairy cattle have a flow that involves a reference population and another population to be selected. Thus, the reference population consists of animals which have, necessarily, accurate information of the trait. This population is used as the genetic basis for predicting the effects of markers.

The determination of the reference population, as its size and its constitution, has great influence on the accuracy of genomic predictions (Hayes et al., 2009; VanRaden et al., 2009). In the case of dairy herds, the constitution is mainly dependent on the composition (bulls and cows) and selective genotyping according to the structure of the response variable. In most countries, only sires, mainly high accuracy were genotyped and included in the reference population (Loberg and Dürr, 2009). For other side, the use of more accurate information implies the use the best animals as reference. However, simulation studies of dairy cattle as Jiménez-Monteiro et al. (2011) concluded that the selection of only females with high estimated breeding values or yield deviations produced suboptimal results. This study showed that the better sampling for females are upper and lower extreme values within the distribution of yield deviations with the usual sampling for males, although these authors have not evaluated this combination with deviations yield of daughters (DYD) for sires.

One of the first steps for genomic selection is generate the response variable

for genomics analysis, which depends on the available sources (individual performance, of the daughters or parents). Different information can be used, from the phenotype itself, as single records, repeated records, the average of the progeny, or even the pseudo-phenotypes as DYD (VanRaden and Wiggans, 1991), EBV and deregressed-EBV (Garrick et al., 2009). For dairy traits, pseudo-phenotypes are preferred because lactations are sex-limited and only females have phenotypes. Among these, DYD are the most-used because besides sires have a larger impact on breeding programs than cows, and their DYDs are more accurate than cow phenotypes (Calus et al., 2009). The deregressed-EBV can be considered a type of deregressed-proof (DRPF) when are used sires with information of the high accuracy in the reference population, because they combine different sources of information about the sires besides the informations of his daughters. Generally the DRPFs are considered equivalent to DYD (Sigurdsson et al., 1995). The deregressed-EBV are used in genomic evaluation of dairy traits when there are cows and bulls in the reference population. However when using any deregressed pseudo-phenotype there is an individual increment disproportionate in response variable that leads to the need to consider the heterogeneity of the residue by statistical models, with weights that range according to the source used to deregress and effects considered (Garrick et al. 2009).

When using genomic evaluations as criteria for pre-selection of animals to progeny test it is possible to reduce spending to prove that animals would have low performance in the test (Hayes et al., 2009), but there is potentially a problem with preselection bias (Patry and Ducrocq, 2011). Dekkers (2006) reported that rates of genetic change can be 3 to 4 times higher with GS than under current progeny test programs, and the savings in logistical costs could be 97% of today's cost. Furthermore, genotyping costs are also likely to decrease over time, which would make the GS easier to administer. Schrooten et al. (2005) reported that genetic progress increased from 19 to 31% compared to progeny testing, when the molecular markers explained 50% of the genetic variance. VanRanden et al. (2009) reported that the

predictive ability for dairy traits using genomic predictions was 50% versus 27% for the traditional PTA. They also reported that gains for proven bulls were highly significant, although smaller than the young bulls because of the higher initial reliability of these bulls. Note that the GS is already applied and showed promising results in many countries, including the USA, Canada, New Zealand, and the Netherlands (Loberg and Dürr, 2009).

Despite these latest innovations in genomics area bring the advantages described above, the molecular informations also demanded an increase in statistical and computational resources, limiting the use of such information in many analyzes including the multi-trait models and test-day models. Although these models are easily applied by replacing the traditional numerator relationship matrix (A) with the genomic relationship matrix (G) (Koivula et al, 2012; Tsuruta et al, 2011), has yet didn't get at a robust analysis of multivariate nature with models that consider different variances for the markers, this being perhaps one of the greatest prospects for optimization of genomic analyzes of the dairy traits.

Parentage correction and Pedigree errors

For a successful and comprehensive evaluation of individuals in any breeding program, correct parentage and pedigree information are essential because pedigree information is a key part of variance component and breeding value estimations. Pedigree error rates in dairy cattle breeds have been estimated to average 10 to 12% average (Banos et al., 2001; Spelman, 2002; Visscher et al., 2002), although reports from the 1970s to the late 1990s estimated values ranging from 5% to about 22% (Christensen et al., 1982; Geldermann et al., 1986; Bovenhuis and Van Arendonk, 1991; Ron et al., 1996). The rapid adoption of micro-satellite parentage testing in the cattle breeding industry probably reduced parentage and pedigree errors, most commercial (grade) cows are not tested. Although error rates may decreased over the years, parentage and pedigree inconsistencies of 10 or 11% can lead to reductions in

genetic gain of 2 to 18% (Banos et al., 2001; Visscher et al., 2002).

Before the advancement in high throughput single nucleotide polymorphism (SNP) data, blood groups (Stormont, 1967) and mini- and micro- satellite (Kashi et al., 1990) were the basic means of inferring parentage. Even though micro-satellite are still used, with the recent availability of SNP markers and with large numbers of sires (Harris and Johnson, 2010; Weigel et al., 2010; Fritz et al., 2013; VanRaden et al., 2013b) and dams (Spelman et al., 2013; VanRaden et al., 2013b) genotyped in the USA, Canada, Australia, New Zealand, Ireland and France among others, parentage and pedigree errors are increasingly identified using SNP genotypes. McClure et al. (2012, 2013) have developed methods to impute microsatellite parentage panels from SNP-based parents panels, which will assist cattle producers as they transition from microsatellite to SNP genotyping for parental verification.

Parentage assignment is a method aimed at excluding individuals ("exclusion principle") from the list of potential parent. This means that, a large number of "potential" sires and dams are examined and only one or a few individuals are retained based on their marker data by using simple segregation rules (Kashi et al., 1990; Hayes, 2011). In addition to marker genotypes, additional accuracy can be achieved if information of birth date and mating records are considered.

Due to the abundant SNP marker information and the shift from micro-satellite to small SNP panels, we give a brief description of how SNP information is used to correct pedigrees and infer potential parents. Initial verification of information (parents) obtained from pedigree are an be done to detect parent-offspring inconsistencies. If parent-offspring errors exceeds a certain defined threshold, then loop through the entire genotype data infer the potential parents. For individuals with no pedigree information, loop through the entire genotype data directly to obtain potential parent.

The algorithm for detecting and inferring parent offspring conflict is based on Mendelian inheritance rules (Calus et al., 2011; Hayes, 2011). This means that, for a bi-allelic SNP, an individual and the prospective parent are both homozygous but for different

alleles “opposing homozygotes”. eg. If an individual has an A|A genotype, the potential parent should carry the “A” allele (A|A or A|B), however if the potential parent has B|B allele for more possible parent offspring conflicts per locus) then they have opposing homozygous genotype. Looping across all SNP genotypes, the sum of all “opposing homozygous” is compared to an empirically determined threshold (determined from on genotype error rate). Furthermore, to avoid picking up monozygotic twins in the pairwise comparison (parent-offspring check), information from birth years could be used.

The empirical thresholds are based on the realized distribution of genotyping errors for all parent-offspring conflict checks. Wiggans et al. (2009), Calus et al. (2011) and Hayes (2011) all reported similar distribution of Mendelian errors for the Illumina 50K SNP panel. Wiggans et al. (2009) and Calus et al. (2011) used >200 SNPs and >250 SNPs, respectively, on a 50K panel to exclude parent-offspring conflicts. However, Hayes (2011) used a stringent threshold of >25 SNPs on 50K SNP panel and >8 SNPs on a 3K SNP panel. Additionally, unpublished results from Gyr (Brazilian *Bos indicus* breed) shows similar distribution when 50K SNP panel was used. The number of markers within the empirical threshold were <58, however <200 SNP markers were used as the threshold. The total parent-offspring conflict observed was about 8% in Gyr and we could infer potential parent for about 2% of these errors. Calus et al. (2011) removed 230 individuals with parent-offspring conflict. Fisher et al. (2009) reported that, about 40 highly polymorphic SNP markers (MAF>0.35) and on-farm information about birth dates and mating periods were needed to correctly assign parentage without any unambiguity.

Using SNP data, (i) recent ancestors errors, (ii) maternal grandsire errors, and (iii) full- and half-sibling errors could also be corrected (Wiggans et al., 2009; Calus et al., 2011; VanRaden et al., 2013a). The reduction in parentage and pedigree errors would go a long way to help reduce the loss in genetic gain (Banos et al., 2001; Visscher et al., 2002) and potentially decrease inbreeding. We conclude that, although

there is a dearth of knowledge on the current pedigree errors detected using SNP data in dairy cattle populations, the supposed reason been, the lack of interest in publishing pedigree errors, however, most publish materials on genomic related research (genomic selection, GWAS, etc.) undertake this key component before performing their analysis.

Principal traits selected in a Dairy Breeding program and their results with genomic selection analyses.

Since 2009, the United States in collaboration with Canada has published genomic evaluations based on BovineSNP50 genotypes. More recently these two countries included Illumina's Bovine3K chip genotypes in their GEBV estimations, increasing enormously the number of genomically evaluated animals (VanRaden et al., 2011a). Many countries like Australia, New Zealand, Germany, Switzerland and many others also implemented GS on their breeding programs and encouraged widespread use of young genomically evaluated bulls (Wiggans, 2011). Hutchison et al. (2014) have recently shown that the heavy use of genomically evaluated young bulls in the US has greatly reduced the generation interval and improved the rate of genetic gain. The average age of sires of Holstein bulls born in 2012 was 2.7 yr younger than those males born in 2006, and 1.3 yr younger for females. This indicates that dairy producers are willing to use semen from young bulls that rank highly rather than use lower-ranking bulls with progeny tests.

One of the most important requirements for GS implementation is the use of a large reference or training population that include animals with both phenotype and genotype information, thus all the traits routinely evaluated on commercial breeding programs are able to have their GEBV estimated. In the United States, more than 30 traits traditionally estimated and related to health, yield, and fertility of dairy cattle have their GEBV available, including net merit, milk yield, protein yield, fat yield, protein percentage, fat percentage, productive life, somatic cell score and daughter pregnant

rate (VanRaden et al. 2009; Weigel et al. 2010).

Previous results of GS from Australia for protein yield, protein percentage, fertility, Australian Profit Ranking and Australian Selection Index, demonstrated that GEBV reliabilities estimated with Bayes A and BLUP methodologies were in a range of 0.14 to 0.48 and 0.18 to 0.44, respectively. The reliabilities of GEBV were considerably greater than traditional EBV estimations even with a small size reference population (approximately 600 animals) (Hayes et al., 2009).

More recent results from Australian Dairy Futures Cooperative Research Centre's demonstrated that the expansion of reference population to 10,000 Holstein and 4000 Jersey cows, could lead to 0.04 to0.08 improvement in the reliability of breeding values depending on the trait (Pryce et al., 2012a).

In a similar study conducted by LIC (Livestock Improvement Corporation) in New Zealand, GEBV for milk production traits, live birth weight, fertility, somatic cell counts and longevity presented reliabilities for young bulls with no daughter information between 0.50 to 0.67, indicating a increase in the rate of genetic trend above than 50% if compared with traditional EBV (Harris et al., 2008).

As milk production becomes increasingly specialized and competitive, selection objectives will need to include traits related to profitability and animal efficiency. To meet this goal, new traits, not traditionally measured by breeding programs, are been evaluated for inclusion in selection programs, such as feed efficiency (De Haas et al., 2012), methane emission (Wall et al., 2010), energy balance (Verbyla et al., 2010), disease resistance (Kirkpatrick et al., 2011; Parker Gaddis et al., 2014), novel fertility traits (Cochran et al., 2013a, 2013b), resistance to heat stress (Dikman et al., 2013), and calf birth weight (Cole et al., 2014). One of the restrictions of the introduction of novel traits on GS is the low accuracies of the estimations due to small to moderate size of the reference population.

Calus et al., 2013 evaluating a novel trait with heritability ranging between 0.05 and 0.30 and a moderate size reference population demonstrate that although the

accuracies are low (0.15 and 0.43 for traits with heritabilities of 0.05 and 0.30, respectively), the selection response could be substantial depending on the heritability and economic value of the new trait and the genetic correlation with the current breeding goal. Accordingly, to achieve accuracies acceptable in dairy cattle breeding programs, the reference population should be larger.

Imputation results on genomic evaluation

The increase in genetic gain and accuracy of prediction using genomic information have been discussed extensively under section 3. Even though the price of genotyping individuals for the implementation was high about a decade ago, the promise of doubling genetic gain at a lower cost than progeny testing (Schaeffer, 2006) was enough incentive to genotype bulls on the then 50K Illumina SNP panel (Matukumalli et al., 2009). However, the extended cost that came along with the requirement of increasing the reference population (training set) and genotyping selection candidate facilitated the need to use alternative SNP panels that were cheaper and preferably efficient for genomic selection. Additionally, the accurate in-silico genotyping (Imputation) of SNP markers in the field of human genetics gave a unique perspective into how genotyping cost could be drastically reduced (Browning and Browning, 2007, 2009; Howie et al., 2009).

Genotype imputation uses population-based linkage disequilibrium (LD), family-based linkage information or a combination of both, to infer genotypes at un-typed marker loci. Population-based imputation algorithms were developed mainly to explore and capture LD information without using a prior family information which might not be available. These methods are very popular in the human genetics field, however it is also heavily used in the field of animal genetics. The most prominent population-based software includes Beagle (Browning and Browning, 2007, 2009), Impute2 (Howie et al., 2009), MaCH (Li et al., 2010), fastphase (Scheet and Stephens, 2006), and PLINK (Purcell et al., 2007). On the other hand, family-based or a combination of population

and family based imputation algorithms have been developed in the field of animal genetics. These algorithms uses *a priori*, the family information and subsequently LD information to infer un-typed markers. Commonly used software includes PHASEBOOK (LinkPHASE and DAGPHASE) (Druet and Georges, 2011), FImpute (Sargolzaei et al., 2012), AlphaImpute (Hickey et al., 2011), Findhap (VanRaden et al., 2011a), and PEDIMPUTE (Nicolazzi et al., 2013).

In dairy cattle breeding programs, to reduce genotyping cost, the Illumina Bovine3K BeadChip (~2,900 SNPs) (Illumina Data Sheet, 2011) and Illumina Bovine7K BeadChip (~6,900 SNPs) (Boichard et al., 2012) were developed. Imputation accuracies from a lower density SNP panel to higher density SNP panels have been pretty accurate. Dassonneville et al. (2011) reported imputation error rate (allelic error rate) of 3.9%, when they imputed from 3K to 50K in French Holstein (reference population = 3,071; validation set = 966) using a combination of Beagle v2.1.3 and DAGPHASE. They also reported 5.5% error rate for Holstein bulls of the three Nordic countries (reference population = 3,058; validation set = 1,086). Increasing the Reference population with bulls from the EuroGenomics consortium reduced error rate to 2.1% in the French Holstein and 4.0% in Holstein bulls from the Nordic countries. Sargolzaei et al. (2011) also reported imputation error rate (imputing 3K to 50K) between 2.2% to 4.1% in three Canadian dairy cattle breeds (Hosltein, Jersey and Brown Swiss). Error rate was lower (between 0.53% to 1.03%) when the 7K SNP panel was used. Khatkar et al. (2012) also reported error rate of about 3.3% for Australian Holstein using Impute2. Recently, Ma et al. (2013) reported allelic error rate of 3.7% in Swedish and Finish Red dairy cattle for imputing 3K to 50K using beagle v3.3. Studies from three Italian dairy cattle breeds (Hosltein, Brown Swiss and Simmental) by Dimauro et al. (2013) showed a lower allelic imputation error rate for imputing 50K from 3K and 7K compared to the results presented above. Error rate were about 10% for 3K and 5% for HD using Beagle v3.3. Others studies with varying subset of the 50K SNP markers shows error rate of about 2% to 8% (Weigel et al., 2010; Khatkar et al., 2012).

The above studies have been done using B*os taurus* breeds, however, initial imputation results from Gyr, an important *Bos indicus* dairy cattle breed of Brazil shows slightly higher allelic error rate (7.0% for 3K and 4.0% for 7K using Beagle v4) than the *Bos taurus* breeds (Boison et al. (2014a), *accepted accepted for the Proceedings of EAAP 2014*). Differences in population structure, number of animals in reference population, choice of imputation algorithms or softwares have been explicitly shown to account for the observed differences across studies. Furthermore, the higher allelic error rate observed for *Bos indicus* breeds might also be due to ascertainment bias of the 3K, 7K and 50K Illumina SNP panels. This results in a small number of markers left on the lower density SNP panel (< 2,000 for 3K and < 5,000 for 7K) as tag SNPs with large inter-marker distances.

To increase accuracies of genomic breeding values more than what was observed with the 50K SNP panel for multi-breed genomic, purebreed and crossbreed evaluations, the BovineHD BeadChip (HD) was introduced. Additionally, the HD SNP panel was to help reduce the ascertainment bias observed for the 50K in *Bos indicus* breeds.

Since most animals were already genotyped on the 50K SNP panel, few but influential sires were suggested to be genotyped on the HD SNP panel and the rest of the population imputed. Goddard and Hayes (2009) suggested an algorithm for efficiently selecting individuals to be genotyped on HD, so that imputation accuracies for the rest of the population would be high. For most breeds, imputing HD genotypes from 50K have been high. Khatkar et al. (2012), Brøndum et al. (2012), Ma et al. (2013), Pausch et al. (2013), Schrooten et al. (2014), and Erbe et al. (2012), all have reported imputation accuracies between nearly zero and 7% when the reference population used in building haplotype library was at least 1/3 of the validation population.

However, accuracies of GEBVs of real data for HD genotypes have been shown to be smaller or sometimes lower than expected in comparison to using the 50K SNP

panel (Erbe et al., 2012; Hayes et al., 2012; Su et al., 2012) regardless of the trait and method of prediction for *Bos taurus* breeds. Initial results in Guzerat; a *Bos indicus* dairy cattle breed of Brazil, shown an increase in accuracies of about 4-12% for milk, fat and protein yield in kg (Boison et al. (2014b), *accepted for the Proceedings of WCGALP 2014*). The differences in the result might be attributed to the right-skewed nature (lot of markers in low frequency due to ascertainment bias) of the minor allele frequency distribution for the 50K observed in most *Bos indicus* breeds. The HD SNP panel, on the other hand, has markers with high MAF.

Interestingly, effect of imputation on "overall" estimate of accuracies of breeding values have been very little for imputing 7K to 50K (Khatkar et al., 2012; Dimauro et al., 2013) and 50K to HD (Khatkar et al., 2012; Su et al., 2012), substantially lower for 3K to 50K (Dassonneville et al., 2011; Dimauro et al., 2013). The great reduction in accuracies of GEBV for the 3K to 50K have been attributed to the lower imputation accuracies observed, compared to the imputation accuracies obtained for 7K to 50K and 50K to HD. Low-density genotypes also are most commonly available for young animals with no phenotypes or cows with few phenotypes available, and the resulting low-accuracy PTA may be more responsible for the small gain, rather than the chip itself. Recently, Druet et al. (2014), have shown with a simulation study that, using imputed sequence data might follow the same trend like the imputed HD (very little loss in accuracy of GEBV). However, they point out in their paper that, effect of QTLs in low frequency might be estimated with lower accuracy than using actual sequence data.

The results from the studies available, suggests that, imputing genotypes from 7K to 50K; 50K to HD is feasible and accurate enough for genomic evaluations.

Other uses of genomic information

There are a number of applications for genomic information other than the prediction of high-reliability breeding values. Perhaps the most prominent recent application is the use of haplotypes in combination with next-generation sequencing

data to identify causal variants associated with recessives. The methodology for identifying recessive haplotypes by searching for a deficit of homozygotes was first described by VanRaden et al. (2011a), and its use in combination with sequence data to identify a causal variant (*APAF1*, associated with the HH1 haplotype) was reported in Adams et al. (2012). Additional details are provided in VanRaden et al. (2012). The US currently tracks 19 recessive haplotypes, and the causal variant for many of those conditions is known (Cole et al., 2013).

While in theory genomic selection should result in lower rates of inbreeding (Daetwyler et al., 2007), that has not proven to be true in practice (e.g., VanRaden et al., 2011b). Sun et al. (2013) have showed that the use of genomic inbreeding coefficients rather than pedigree inbreeding in mating programs results in decreases in expected progeny inbreeding, and the economic value of using genomic relationships is >$3 million per year for US Holsteins when applied to all genotyped females. These results are consistent with the work of Pryce et al. (2012b), who also found that it is beneficial to consider genomic inbreeding when allocating mates. However, Cole and VanRaden (2011) showed that the best chromosomal genotypes generally consist of two copies of the same haplotype, even after adjustment for inbreeding, underscoring the tension between strategies that ensure maximal rates of genetic gain versus those that try to balance selection response against the need to maintain genetic diversity in the population.

References

Adams, H.A., Sonstegard, T., VanRaden, P.M., Null, D.J., Van Tassell, C., Lewin, H., 2012. Identification of a nonsense mutation in APAF1 that is causal for a decrease in reproductive efficiency in dairy cattle. Proc. Plant Anim. Genome XX Conf., abstr. P0555.

Banos, G., Wiggans, G.R., and Powell, R.L. , 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. J. Dairy Sci. 84,

2523–2529.

Bohmanova, J., Sargolzaei, M., Schenkel, F., 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics 11, 421-432.

Boichard, D., Chung, H., Dassonneville, R., David, X., Eggen, A., Fritz, S., Gietzen, K.J., Hayes, B.J., Lawley, C.T., Sonstegard, T.S., et al. , 2012. Design of a bovine low-density SNP array optimized for imputation. PLoS One 7, e34130.

Boison, S.A., Santos, D., Utsunomiya, A., Garcia, F., Verneque, R., Silva, M.V.B., and Sölkner, J. , 2014a. Genotype imputation in Gir (Bos indicus): comparing different commercially available SNP chips. In Proceeding of EAAP (Book of Abstract), p. 254.

Boison, S.A., Santos, D.J. de A., Garcia, J.F., Sölkner, J., Peixoto, M.G.C.D., and Silva, M.V.G.B. da , 2014b. Genomic Evaluation Using 50K and Imputed HD Genotypes in Guzera (Bos indicus) Breed. In Proceedings of the WCGALP, (Vancouver, Canada), pp. 3908–3911.

Bovenhuis, H., and Van Arendonk, J. a , 1991. Estimation of milk protein gene frequencies in crossbred cattle by maximum likelihood. J. Dairy Sci. 74, 2728–2736.

Browning, S.R., Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81, 1084–1097.

Browning, B.L., and Browning, S.R., 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84, 210–223.

Browning, S.R., Browning, B.L., 2010. High-resolution detection of identity by descent in unrelated individuals. Amer. J. Human Genet. 86, 526-539.

Brøndum, R.F., Ma, P., Lund, M.S., and Su, G. , 2012. Short communication: genotype imputation within and across Nordic cattle breeds. J. Dairy Sci. 95, 6795–6800.

Calus, M.P.L., 2009. Genomic breeding value prediction: methods and procedures. Animal 4, 157-164.

Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W., Veerkamp, R.F., 2008. Accuracy of

genomic selection using different methods to define haplotypes. Genetics 178,553-561.

Calus, M.P.L., Mulder, H.A., Bastiaansen, J.W.M., 2011. Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genet. Sel. Evol. 43, 34.

Calus, M.P.L., de Haas, Y., Pszczola, M., Veerkamp, R.F., 2012. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. Animal 7, 183–191.

Christensen, L.G., Madsen, P., Petersen, J., 1982. The influence of incorrect sire identification on the estimates of genetic parameters and breeding values. In Proceedings of the WCGALP, (Madrid, Spain), pp. 200–208.

Cochran, S.D., Cole, J.B., Null, D.J., Hansen, P.J., 2013a. Discovery of single nucleotide ploymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genet. 14, 49.

Cochran, S.D., Cole, J.B., Null, D.J., Hansen, P.J., 2013b. Single nucleotide polymorphisms in candidate genes associated with fertilizing ability of sperm and subsequent embryonic development in cattle. Biol. Reprod. 89, 69.

Cole, J.B., VanRaden, P.M., 2011. Use of haplotyes to estimate Mendelian sampling effects and selection limits. J. Anim. Breed. Genet. 128, 448-455.

Cole, J.B., VanRaden, P.M., Null, D.J., Hutchison, J.L., Cooper, T.A., 2013. AIPL Research Report GENOMIC3: Haplotype tests for recessive disorders that affect fertility and other traits. Accessed May 8, 2014. http://aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html.

Cole, J.B., VanRaden, P.M., O'Connell, J.R., Van Tassell, C.P., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., Wiggans, G.R., 2009. Distribution and location of genetic effects for dairy traits. J. Dairy Sci. 92, 2931–2946.

Cole, J.B., Waurich, B., Wensch-Dorendorf, M., Bickhart, D.M., Swalve, H.H., 2014. A genome-wide association study of calf birth weight in Holstein cattle using single nucleotide polymorphisms and phenotypes predicted from auxiliary traits. J. Dairy Sci. 97, 3156–3172.

Daetwyler, H.D., Villanueva, B., Bijma, P., Woolliams, J.A., 2007. Inbreeding in genome- wide selection. J. Anim. Breed. Genet. 124, 369-376.

Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J., Lander, E.S., 2001. High-resolution haplotype structure in the human genome. Nature Genet. 29, 229-232.

Dassonneville, R., Brøndum, R.F., Druet, T., Fritz, S., Guillaume, F., Guldbrandtsen, B., Lund, M.S., Ducrocq, V., Su, G., 2011. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. J. Dairy Sci. 94, 3679–3686.

De Haas, Y., Calus, M.P.L., Veerkamp, R.F., Wall, E., Coffey, M.P., Daetwyler, H.D., Hayes, B.J., Pryce, J.E., 2012. Improved accuracy of genomic prediction for dry matter intake of dairy cattle from combined European and Australian data sets. J. Dairy Sci. 95, 6103-6112.

de Roos, A.P.W., Hayes, B.J., Spelman, R.J., Goddard, M.E., 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179, 1503-1512.

Dekkers, J.C.M., van Arendonk, J. A. M., 1998. Optimum selection for quantitative traits with information on an identified locus in outbred populations. Genet. Res. 71, 257-275.

Dekkers, J.C., 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Animal Sci. 82, 313–328.

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.

Dikmen, S., Cole, J.B., Null, D.J., Hansen, P.J., 2013. Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. PLoS ONE 8, e69202.

Dimauro, C., Cellesi, M., Gaspa, G., Ajmone-Marsan, P., Steri, R., Marras, G., Macciotta, N.P., 2013. Use of partial least squares regression to impute SNP genotypes in Italian Cattle breeds. Genet. Sel. Evol. 45, 15.

Druet, T., Georges, M., 2010. A hidden markov model combining linkage and linkage

disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. Genetics 184, 789–798.

Druet, T., Macleod, I.M., Hayes, B.J., 2014. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. Heredity (Edinb). 112, 39–47.

Elsik, C.G., Tellam, R.L., Worley, K.C., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324, 522–528.

Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., Reich, C.M., Mason, B.A., Goddard, M.E., 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95, 4114–4129.

Falconer D.S., Mackay T.C., 1997. Introduction to Quantitative Genetics. 4th Edition. John Wiley & Sons, Hoboken, NJ.

Fisher, P.J., Malthus, B., Walker, M.C., Corbett, G., Spelman, R.J., 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. J. Dairy Sci. 92, 369–374.

Fritz, S., Capitan, A., Djari, A., Rodriguez, S.C., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., et al., 2013. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. PLoS One 8, e65550.

Garrick, D.J., Taylor, J.F., Fernando, R.L., 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41, 1-8.

Geldermann, H., Pieper, U., Weber, W.E., 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. J. Anim. Sci. 63, 1759–1768.

Griffiths, A.J.F., Wessler, S.R., Lewontin R.C., Carroll, S.B., 2007. Introduction to Genetic Analysis. 9th Edition. W.H. Freeman & Company, New York, NY.

Harris, B.L., Johnson, D.L., 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. 93, 1243–1252.

Harris, B.L., Johnson, D.L., Spelman, R.J., 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Proc. Interbull Meeting, Niagara Falls, Canada, 2008.

Hayes, B.J., Lien, S., Nilsen, H., Olsen, H.G., Berg, P., MacEachern, S., Potter, S., Meuwissen, T.H.E., 2008. The origin of selection signatures on bovine chromosome 6. Anim. Genet. 39, 105-116.

Hayes, B.J., Visscher, P.M., Mcpartlan, H.C.E., Goddard, M. E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13, 635–643

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

Hayes, B.J. , 2011. Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. J. Dairy Sci. 94, 2114–2117.

Hayes, B.J., Daetwyler, H.D., Bowman, P., Moser, G., Tier, B., Crump, R., Khatkar, M., Raadsma, H.W., Goddard, M.E., 2012. Accuracy of genomic selection: comparing theory and results. Deliv. Genomics to Ind. 34–37.

Hazel, L.N., Lush, J.L., 1943. The efficiency of three methods of selection. J. Heredity 33, 393-399.

Henderson, C.R., 1949. Estimation of changes in herd environment. J. Dairy Sci. 32, 709.

Hickey, J.M., Kinghorn, B.P., Tier, B., Wilson, J.F., Dunstan, N., van der Werf, J.H.J., 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genet. Sel. Evol. 43, 12.

Howie, B.N., Donnelly, P., Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS

Genet. 5, e1000529.

Hutchison, J.L., Cole, J.B., Bickhart, D.M., 2014. Short communication: Use of young bulls in the United States. J. Dairy Sci. 97, 3213–3220.

Illumina Inc. 2011. GoldenGate Bovine3K Genotyping BeadChip. Accessed May 8, 2014.

http://www.illumina.com/Documents/products/datasheets/datasheet_bovine3K.pdf.

Jiménez-Montero, J.A., González-Recio, O., Alenda, R., 2011. Genotyping strategies for genomic selection in small dairy cattle populations. Animal 6, 1216-1224.

Kashi, Y., Hallerman, E., Soller, M., 1990. Marker assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. 51, 63-74.

Kashi, Y., Lipkin, E., Darvasi, A., Nave, A., Gruenbaum, Y., Beckmann, J.S., Soller, M., 1990. Parentage identification in the bovine using “deoxyribonucleic acid fingerprints”. J. Dairy Sci. 73, 3306–3311.

Kaupe, B., Winter, A., Fries, R., Erhardt, G., 2004. DGAT1 polymorphism in Bos indicus and Bos taurus cattle breeds. J. Dairy Res. 71, 182–187.

Khatkar, M.S., Zenger, K.R., Hobbs, M., Hawken, R.J., Cavanagh, J.A.L., Barris, W., McClintock, A.E., McClintock, S., Thomson, P.C., Tier, B., Nicholas, F.W., Raadsma, H.W., 2007. A primary assembly of a bovine haplotype block map based on a 15,036 single nucleotide polymorphism panel genotyped in Holstein Friesian cattle. Genetics 176, 763–72.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A., Barris W., Schnabel R.D., Taylor J.F., Raadsma H.W., 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics 9, 187.

Khatkar, M.S., Moser, G., Hayes, B.J., Raadsma, H.W., 2012. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. BMC Genomics 13, 538.

Kirkpatrick, B.W., Shi, X., Shook, G.E., Collins, M.T., 2011. Whole-genome association analysis of susceptibility to paratuberculosis in Holstein cattle. Anim. Genet. 42, 149–160.

Koivula, M., Strandén, I., Pösö, J., Aamand, G. P., Mäntysaari, E.A., 2012. Single step genomic evaluations for the Nordic Red dairy cattle test day data. Interbull Bull. 46, 28 - 31.

Li, Y., Willer, C.J., Ding, J., Scheet, P., Abecasis, G.R., 2010. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. 34, 816–834.

Loberg, A., Dürr, J.W., 2009. Interbull survey on the use of genomic information. Interbull Bull. 39, 3-14.

Ma, P., Brøndum, R.F., Zhang, Q., Lund, M.S., Su, G., 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. J. Dairy Sci. 96, 4666–4677.

Mackinnon, M.J., Georges, M.A.J., 1998. Marker-assisted preselection of young dairy sires prior to progeny-testing. Livest. Prod. Sci. 54, 229-250.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., Connell, J.O., Moore, S.S., Smith, T.P.L., Sonstegard, T.S., et al., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, 1–12.

McClure, M., Sonstegard, T., Wiggans, G., Van Tassell, C. P., 2012. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. Front. Genet. 3:140.

McClure, M.C., Sonstegard, T.S., Wiggans, G.R., Van Eenennaam, A.L., Weber, K.L., Penedo, C.T., Berry, D.P., Flynn, J., Garcia, J.F., Carmo, A.S., Regitano, L.C.A., Albuquerque, M., Silva, M.V.G.B., Machado, M.A., Coffey, M., Moore, K., Boscher, M-

Y.., Genestout, L., Mazza, R., Taylor, J.F., Schnabel, R.D., Simpson, B., Marques, E., McEwan, J.C., Cromie, A., Coutinho, L.L., Kuehn, L.A., Keele, J.W., Piper, E.K., Cook, J., Williams, R., Bovine HapMap Consortium, Van Tassell, C.P., 2013. Imputation of microsatellite alleles from dense SNP genotypes for parentage verification across multiple Bos taurus and Bos indicus breeds. Front. Genet. 4:176.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Wouter Coppieters, W., Crews, D., Dias Neto, E., Gill, C.A., Gao, C., Mannen, H., Stothard, P., Wang, Z., Van Tassell, C.P., Williams, J.L., Taylor, J.F., Stephen, S., Moore, S.S. Whole genome linkage disequilibrium maps in cattle. BMC Genetics 8, 74.

Meuwissen, T.H.E., Goddard, M.E., 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28, 161-176.

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

Meuwissen, T.H.E., 2007. Genomic selection: marker assisted selection on genome-wide scale. J. Anim. Breed. Genet. 124, 321-322.

Muir, W.M., 2007. Comparison of genomic and traditional BLUP – estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124, 342-355.

Nejati-Javaremi, A., Smith, C., Gibson, J.P., 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Animal Sci. 75, 1738–1745.

Nicolazzi, E.L., Biffani, S., Jansen, G., 2013. Short communication: imputing genotypes using PedImpute fast algorithm combining pedigree and population information. J. Dairy Sci. 96, 2649–2653.

Parker Gaddis, K.L., Cole, J.B., Clay, J.S., Maltecca, C., 2014. Genomic selection for producer-recorded health event data in US dairy cattle. J. Dairy Sci. 97, 3190–3199.

Patry, C., Ducrocq, V., 2011. Evidence of biases in genetic evaluations due to genomic

preselection in dairy cattle. J. Dairy Sci. 94, 1011-1020.

Pausch, H., Aigner, B., Emmerling, R., Edel, C., Götz, K.-U., Fries, R., 2013. Imputation of high-density genotypes in the Fleckvieh cattle population. Genet. Sel. Evol. 45, 3.

Pryce, J.E., Hayes, B.J., Goddard, M.E., 2012a. Genotyping dairy females can improve the reliability of genomic selection for young bulls and heifers and provide farmers with new management tools. Proceedings of ICAR Congress: Cork, Ireland.

Pryce, J.E., Hayes, B.J., Goddard, M.E., 2012b. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. J. Dairy Sci. 95, 377–388.

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

Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R., Simianer H., 2010. The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet. 41, 346–356.

Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A., Barbujani, G., 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res. 12, 602-612.

Ron, M., Blanc, Y., Band, M., Ezra, E., Weller, J.I., 1996. Misidentification rate in the Israeli dairy cattle population and Its implications for genetic improvement. J. Dairy Sci. 79, 676–681.

Santos, D.J.A., Utsunomiya, A.T.H., Tonhati, H., Peixoto, M.G.C.D., Panetto, J.C.C.,

Sargolzaei, M., Schenkel, F., Chesnais, J., 2011. Accuracy of imputed 50K genotypes from 3K and 6k chips using FImpute version 2 (ON, Canada).

Sargolzaei, M., Chesnais, J.P., Schenkel, F., 2012. Efficient combined family and population imputation in large data sets. Open Ind. Sess. Oct. 30, 2012 1–10.

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

Scheet, P., Stephens, M., 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78, 629–644.

Schrooten, C., Dassonneville, R., Ducrocq, V., Brøndum, R.F., Lund, M.S., Chen, J., Liu, Z., González-Recio, O., Pena, J., Druet, T., 2014. Error rate for imputation from the Illumina BovineSNP50 chip to the Illumina BovineHD chip. Genet. Sel. Evol. 46, 10.

Sigurdsson, A., Banos, G., 1995. Dependent variables in International sire evaluations. Acta Agric. Scand. 4, 209-217.

Silva, M.V.G.B., Santos D.J.A., Utsunomiya, A.T.H., Verneque, R.S., Machado, M.A., Panetto, J.C.C., 2013. Persistência da fase para principais raças zebuínas orientais criadas no Brasil. XXIII Reunión de la ALPA, Havana, Cuba.

Smith, C., 1967. Improvement of metric traits through specific genetic loci. Anim. Prod. 9:349–358.

Spelman, R.J. , 2002. Utilisation of molecular information in dairy cattle breeding. In Proceedings of the WCGALP, (Montpellier, France), pp. 20–25.

Spelman, R.J., Hayes, B.J., Berry, D.P., 2013. Use of molecular technologies for the advancement of animal breeding: genomic selection in dairy cattle populations in Australia, Ireland and New Zealand. Anim. Prod. Sci. 53, 869–875.

Stock, K.F., Reents, R.. 2013. Genomic selection: Status in different species and challenges for breeding. Reprod Domest Anim. 48 Suppl 1, 2-10.

Stormont, C., 1967. Contribution of blood typing to dairy science progress. J. Dairy Sci. 50, 253–260.

Su, G., Brøndum, R.F., Ma, P., Guldbrandtsen, B., Aamand, G.P., Lund, M.S., 2012. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. J. Dairy Sci. 95, 4657–4665.

Sun, C., VanRaden, P.M., O'Connell, J.R., Weigel, K.A., and Gianola, D., 2013. Mating programs including genomic relationships and dominance effects. J. Dairy Sci. 96, 8014–8023.

Tsuruta, S., Aguilar, I., Misztal, I., Lawlor, T.J., 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94, 4198-4204.

VanRaden, P.M., Wiggans, G.R., 1991. Derivation, calculation and use of national animal model information. J. Dairy Sci. 74, 2737-2746.

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

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

VanRaden, P.M., Olson, K.M., Null, D.J., Hutchison, J.L., 2011a. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94, 6153–6161.

VanRaden, P.M., Olson, K.M., Wiggans, G.R., Cole, J.B., Tooker, M.E., 2011b. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 94, 5673–5680.

VanRaden, P.M., Null, D.J., Sonstegard, T.S., Adams, H.A., Van Tassell, C.P., Olson, K.M., 2012. Fine mapping and discovery of recessive mutations that cause abortions in dairy cattle. J. Dairy Sci. 95(Suppl. 2), ii–iii(abstr. LB6).

VanRaden, P.M., Cooper, T.A., Wiggans, G.R., O'Connell, J.R., Bacheller, L.R., 2013a. Confirmation and discovery of maternal grandsires and great-grandsires in dairy cattle. J. Dairy Sci. 96, 1874–1879.

VanRaden, P.M., Null, D.J., Sargolzaei, M., Wiggans, G.R., Tooker, M.E., Cole, J.B., Sonstegard, T.S., Connor, E.E., Winters, M., van Kaam, J.B.C.H.M., et al., 2013b. Genomic imputation and evaluation using high-density Holstein genotypes. J. Dairy Sci. 96, 668–678.

Van Tassell, C.P., Van Vleck, L.D., 1991. Estimates of genetic selection differentials and generation intervals for four paths of selection. J. Dairy Sci. 74, 1078–1086.

Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Meth. 5, 247-252.

Verbyla, K.L., Calus, M.P.L., Mulder, H.A., de Haas, Y., Veerkamp, R.F. 2010, Predicting energy balance for dairy cows using high-density single nucleotide polymorphism information. J. Dairy Sci. 93, 2757–2764.

Visscher, P.M., Woolliams, J.A., Smith, D., Williams, J.L. , 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J. Dairy Sci. 85, 2368–2375.

Wall, E., Simm, G., Moran, D., 2010. Developing breeding schemes to assist mitigation of greenhouse gas emissions. Animal 4, 366–376.

Weigel, K.A., de los Campos, G., Vazquez, A., Van Tassell, C.P., Rosa, G.J.M., Gianola, D., O'Connell, J.R., VanRaden, P.M., Wiggans, G.R., 2010. Genomic selection and its effects on dairy cattle breeding programs. Proceedings of the Ninth World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig. 119:8.

Weigel, K.A., Van Tassell, C.P., O' Connel, J.R., VanRaden, P.M., Wiggans, G.R., 2010. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. J. Dairy Sci. 93, 2229-2238.

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Schenkel, F.S., Van Tassell, C.P., 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92, 3431–3436.

Wiggans GR, VanRaden PM , Cooper TA., 2011. The genomic evaluation system in the United States: Past, present, future J. Dairy Sci. 94, 3202–3211.

Yan, J., Shah, T., Warburton, M.L., Buckler, E.S., McMullen, M.D., Crouch, J., 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4, 1-14.

• Genomic methodological basis: linkage disequilibrium, haplotype, phase persistence and imputation methodologies

 Traditional marker-assisted selection and genome selection evaluation: efficiency, profitability, use of (pseudo-) phenotype

 Principal traits selected in a Dairy Breeding program and their results with genomic selection analyses.