https://doi.org/10.17221/83/2020-CJAS

The use of genomic data and imputation methods in dairy cattle breeding

Anita Klímová[1,2], Eva Kašná[2], Karolína Machová[1], Michaela Brzáková[2], Josef Přibyl[2], Luboš Vostrý[1]*

[1]*Department of Genetics and Breeding, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague – Suchdol, Czech Republic*
[2]*Department of Genetics and Breeding of Farm Animals, Institute of Animal Science, Prague – Uhříněves, Czech Republic*

**Corresponding author: kranjcevicova.anita@vuzv.cz*

Citation: Klímová A., Kašná E., Machová K., Brzáková M., Přibyl J., Vostrý L. (2020): The use of genomic data and imputation methods in dairy cattle breeding. Czech J. Anim. Sci., 65: 445–453.

Abstract: The inclusion of animal genotype data has contributed to the development of genomic selection. Animals are selected not only based on pedigree and phenotypic data but also on the basis of information about their genotypes. Genomic information helps to increase the accuracy of selection of young animals and thus enables a reduction of the generation interval. Obtaining information about genotypes in the form of SNPs (single nucleotide polymorphisms) has led to the development of new chips for genotyping. Several methods of genomic comparison have been developed as a result. One of the methods is data imputation, which allows the missing SNPs to be calculated using low-density chips to high-density chips. Through imputations, it is possible to combine information from diverse sets of chips and thus obtain more information about genotypes at a lower cost. Increasing the amount of data helps increase the reliability of predicting genomic breeding values. Imputation methods are increasingly used in genome-wide association studies. When classical genotyping and genome-wide sequencing data are combined, this option helps to increase the chances of identifying loci that are associated with economically significant traits.

Keywords: genomic breeding values; genomic selection; genotyping; microarray; SNP

Introduction

In a conventional animal evaluation based only on pedigree and phenotype information, the accuracy of the average parental breeding value (parent average) is too low to support intensive selection of bulls at birth. Thus, bull selection is usually done around the age of five years, when the phenotypes of the daughters are already known (Jenko et al. 2017). The potential goal of genomic selection is to increase the accuracy of breeding values at an early age. To achieve this goal, it is necessary to genotype enough individuals that already have a phenotype record or have a phenotype record for their offspring (Meuwissen et al. 2016). If the reliability of the genomic breeding value is sufficiently high, then the age selection threshold for the parents of the future generation can be reduced, and the generation interval can be shortened. Shortening the generation interval could lead

Supported by the Ministry of Agriculture of the Czech Republic (Projects No. QK1810253 and MZE-RO0718), by the Ministry of Education Youth and Sports (Project No. LTAUSA19117) and by the Czech University of Life Sciences Prague (Project No. SV19-07-21360).

https://doi.org/10.17221/83/2020-CJAS

to a doubling of genetic gain over selection based on conventional breeding values (Schaeffer 2006).

Genomic selection

Genomic selection has shaped modern breeding programmes for dairy cattle and has contributed significantly to increasing genetic gain for a variety of economically significant traits (Hayes et al. 2009a; VanRaden et al. 2009). An increase in genetic gain can arise from a shorter generation interval, increased intensity of selection, and greater precision in the selection of animals for breeding (Schaeffer 2006). The most significant advantage of incorporating genomics into cattle breeding is mainly for low hereditary traits which, after the inclusion of genomic data, increase the reliability of the prediction of breeding values (Garcia-Ruiz et al. 2016; Wiggans et al. 2017). As molecular methods were evolving, it was possible to include not only performance data and pedigrees of animals but also genome data in animal evaluation. Single nucleotide polymorphisms (SNPs) can be used to estimate the genetic regression coefficients of SNP markers for individual properties and to determine the relationships between individual animals. Therefore, it is possible to calculate the genomic breeding value (GEBV) of young animals with higher reliability than was the case with the standard breeding value prediction. These young individuals can be subsequently used in breeding though no test mating has occurred (Schaeffer 2006). Therefore, biochips have been developed to allow farmers to obtain information about the genotypes of animals. Chips exist in varying densities, with lower density chips offered to reduce the cost. Imputation of SNP markers is one of the key strategies to generate a common high density set of SNPs across all animals (Zhang and Druet 2010). Single nucleotide polymorphism information is most often used to predict genomic values. By including genotypes, it is possible to compile a matrix of realised relationships between individuals, based on the proportions of the genome that are identical by descent between two individuals. Replacing the relationship matrix with the realised relationship leads to an increase in the accuracy of breeding values, especially in individuals who do not have any phenotypic data (Hayes et al. 2009b). In pedigree matrix A, relationships are the expected sharing of the genome of two individuals. However, the realised relationships often differ from this expectation because the size of the genome is finite, and the loci are connected (VanRaden 2008). These deviations are the reason for increasing the accuracy of genomic breeding values. These variations tend to be small, but the smaller the relationship between individuals, the higher the coefficient of variation for their actual relationship. The genomic relationship matrix is thus an estimate of the correct proportion of the genome shared between two individuals. The G matrix is independent of the pedigree and is estimated based on the SNP (Hayes et al. 2009b). Statistical methods for GEBV prediction fall into three basic categories. The first is BLUP (best linear unbiased prediction) models, which typically use SNPs to extend or replace family tree data to create a relationship matrix. Another category is Bayesian regression models, which assume that the effects of SNP markers are sampled from the distribution or a mixture of several distributions. The last category is machine learning algorithms that allow highly flexible modelling of associations between SNP markers and the desired trait (Weigel 2017). The selection of animals is based on relationships in the genomic relationship matrix, which can be combined with phenotypic information of non-genotyped animals using the BLUP model. The most commonly used method is the multi-step (ms) method GBLUP, which combines the previously mentioned sources of information. The main problem with msGBLUP is that it does not consider genomic preselection in the calculation of evaluations based only on phenotypic data. In some cases, this method is not optimal due to its complexity. It is possible to use a more straightforward computational method called the single-step method or ssGBLUP. This method uses a merged relationship matrix H, which is formed by combining matrix G and A in one step (Misztal et al. 2013). Matrix G represents the genomic relationship matrix for genotyped animals and matrix A represents the pedigree relationship matrix. This procedure allows the use of genetic markers to assess the entire population (Misztal et al. 2009; Bauer et al. 2015). In routine estimates of genomic breeding values, methods based on BLUP models are the most commonly used, namely, the multi-step method of estimating GEBV-GBLUP and the single-step method of estimating GEBV-ssGBLUP (Misztal et al. 2009; Aguilar et al. 2010; Christensen and Lund 2010).

https://doi.org/10.17221/83/2020-CJAS

Development of genotyping

The incorporation of genomic selection into the breeding of dairy cattle has caused a revolution that has resulted in the development of new genotyping technologies. In 2000, a technology was developed testing a large number of SNPs at a relatively low cost (Wiggans et al. 2016). As early as in 2005, the MegAllele genotyping bovine 10 000-SNP chip was offered to the public by ParAllele Bioscience, which is now part of Affymetrix, which was designed to detect 10 410 SNP markers (Khatkar et al. 2007). Despite the high public interest in this product, this chip was not suitable for genomic selection purposes, because it did not have an appropriate distribution of SNP markers across the genome. It was necessary to create a better panel for SNP mapping based on informative markers in each block of binding imbalance across the genome (Hayes et al. 2006). For capturing all linkage disequilibrium blocks in Holstein cattle, it would be necessary to include 250 000 markers (Khatkar et al. 2007). In 2006, the United States Department of Agriculture funded two projects to support the further development of a suitable genotyping panel. The first project focused on creating an atlas of cow genes using the Solexa sequencing platform, which is now part of Illumina. The second project was aimed at the development and testing of genotyping suitable for use in genomic selection. Based on the results of these two projects and two research projects from the University of Missouri, University of Alberta, and iBMC, a consortium was established to develop a panel of 60 000 markers for genomic selection and association studies in cattle in cooperation with Illumina (Wiggans et al. 2016). More than 58 000 SNPs were designed for the BovineSNP50 BeadChip (illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf). This chip was officially released commercially in December 2007 and it offered detection of 54 001 SNPs. In July 2010, Illumina released two new chips: the Bovine3K detecting 2 900 SNPs (illumina.com/Documents/products/datasheets/datasheet_bovine3k.pdf) and the BovineHD chip designed to detect 777 962 SNPs (illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf). In 2011, Illumina released a BovineLD chip for detecting 6 909 SNPs, which was a vast improvement over the Bovine3K chip. This was due to the ability to analyse large amounts of data while using the same chemistry as in the BovineSNP50 BeadChip (Wiggans et al. 2016). The development of this technology has dramatically increased the demand for chips of varying density (Nicolazzi et al. 2014). The BovineSNP50 BeadChip is undoubtedly one of the most widely used commercial chips for cattle genotyping from the Illumina company. The third version of this chip is currently available, which offers detection of 53 714 SNPs that are evenly distributed throughout the cattle genome (illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf). Other major producers of chips for genotyping of cattle are Zoetis, Neogen and Thermo Fisher. Zoetis offers a 50K chip for cattle genotyping and Neogen offers GGP Bovine 50K (50 000 SNP) and GGP Bovine 150K (150 000 SNP) chips (www.neogen.com/solutions/genomic-profiles). The above-mentioned chips are based on the Illumina BovineSNP50 BeadChip. Both companies are significantly involved in genotyping the US cattle population. Another of these companies, Thermo Fisher, offers Affymetrix platform chips for cattle genotyping: the Axiom Bovine Genotyping array (> 67 000 SNP) and the Axiom Genome-Wide BOS array (648 855 SNP) (assets.thermofisher.com/TFS-Assets/GSD/brochures/axiom-genotyping-arrays-agrigenomics-brochure.pdf).

Genotyping methods

The first method of genotyping is the microarray method, which is the basis of commercial chips. Analysis and evaluation of the DNA chip is based on the binding of the oligonucleotide probe by covalent binding to the plate. The DNA sample to be evaluated must first be purified by electrophoresis or PCR and then converted to cDNA by reverse transcription. Subsequently, PCR amplification is performed to ensure a sufficient sample volume. In the next step, several molecules of a fluorescent substance are attached to each molecule which can then be detected. The result of this process is a sample containing single-stranded DNA that is labelled with a detectable substance. After contact with the chip, the sample is hybridized with complementary probes and subsequently flushed to cleanse molecules that did not attach to the probes with sufficient hydrogen bridges. The sample is then evaluated by laser. The dyes present on the sample emit the light of a specific wavelength. The number of complementary mol-

https://doi.org/10.17221/83/2020-CJAS

ecules in the sample can be determined by the light emitted (Starkley and Elaswarapu 2010). There are two different genotyping technologies provided by Illumina and Affymetrix. There is also an increasing interest in custom-made chips that are not commercially available or require consent to be used. The advantage of these chips is primarily the inclusion of additional SNP mutations depending on the population (mutations for various diseases, etc.) (Nicolazzi et al. 2014). In the Affymetrix chip, DNA samples are digested with restriction enzymes into different lengths. Adapters are connected to the ends of the sections. Sections of 250–1 000 base pairs are amplified, fluorescently labelled, and hybridized to the probes. The chips are then evaluated by laser, and a specialized computer program is used to determine the genotype. In the Illumina chip, DNA samples are amplified and then digested into smaller parts. These parts hybridize to specific "beads" of the chip, each carrying two probes. Thus, it is possible to genotype both alleles at the SNP locus simultaneously. DNA is ligated at the labelled base site upstream of the SNP locus. Elongated samples are stained and evaluated by laser, and genotypes are determined in a specialized computational program (Walker and Siminovitch 2007).

Genotyping by sequencing (GBS) is a new technology for genotyping that uses restriction endonucleases to cleave genomic DNA. Thus, it uses the sequential cleavage of DNA fragments and is an efficient method for the discovery of SNP markers (Elshire et al. 2011). Genotyping by sequencing methods do not consider previous information about the reference genome of the studied species (Robledo et al. 2017). In addition to the discovery of new SNP markers, GBS techniques have revolutionized evolutionary genomics (Andrews et al. 2016). Genotyping by sequencing techniques have a broad weakness in missing genotypes, and therefore data imputation is necessary for using GBS (Wang et al. 2020).

Imputation methods

Genotype imputation means adding missing SNP markers to increase the amount of genomic information and reduce the cost of livestock genotyping. Imputation is mostly used for the enlargement of genomic data of animals that are genotyped on low-density chips. For this, genomic information on animals from the reference population which are genotyped in a higher-density panel is important (Sargolzaei et al. 2014).

Imputation methods used in livestock breeding are generally divided into two groups: linkage disequilibrium (LD) techniques using Impute2 (Howie et al. 2009), Beagle (Browning and Browning 2009), Mach (Li et al. 2010) and pedigree and segregation-based techniques (LE) or a combination of pedigree, segregation and population information which typically include AlphaImpute (Hickey et al. 2012), Findhap (VanRaden et al. 2011), DAGPHASE (Druet and Georges 2010), FImpute (Sargolzaei et al. 2008; 2014) and PedImpute (Nicolazzi et al. 2013). Population-based imputation methods provide certain benefits. One of them is that information about pedigrees is not very often available or is very often incomplete. These methods provide more accurate imputation for common SNP variants than pedigree imputation (Cheung et al. 2013). Another disadvantage may be that some methods based on pedigree data require the availability of dense genotypes for all close ancestors (Hickey et al. 2012). All existing imputation methods are essentially based on searching similar haplotypes of the observed genotype of animals on the reference panel (Howie et al. 2009). The most widely used are Beagle (Browning and Browning 2009) based on the Hidden Markov models (HMM) and FImpute (Sargolzaei et al. 2014) which imputes based on the Overlapping sliding windows model (OSW). The accuracy of imputation is influenced by several factors such as the number and composition of individuals in the reference group, the effective population size, the allele frequencies and the differences between the densities of the reference and imputed genotypes (Sargolzaei et al. 2014).

A significant factor is also the quality of the genotype, which is evaluated according to the call rate. The genotype call rate refers to the fraction of called SNPs from the total number of SNPs of a given chip. A quality genotype is determined if the call rate reaches a value between 90% and 95%. Accurate input genotypes are an essential condition for correct imputation. Errors in genotypes affect phasing and imputation, resulting in the genotype of the offspring not corresponding to the parental haplotypes (Cooper et al. 2013).

https://doi.org/10.17221/83/2020-CJAS

Beagle and Hidden Markov model

Beagle is imputation software that was created by Brian Browning at the University of Washington. It was created to work with genotypes, their phasing and imputation of missing markers. Beagle uses the localized haplotype cluster model in the imputation model, which is referred to as the Hidden Markov model (Browning and Browning 2013). The model assumes for each marker a variable number of hidden states representing local clusters. Each cluster represents only one possible allele. By detecting the length of IBDs (identical by descent) shared among individuals, the number of hidden states – clusters is lower, thus speeding up the calculations (Wang et al. 2016). Identification of IBD segments proceeds in two steps. In the first step, candidate IBD segments are detected using linear approaches, and in the second step, the IBD segments are refined to obtain identical haplotypes based on probability analysis (Browning and Browning 2013).

FImpute and Overlapping sliding windows model

FImpute is an efficient tool based on the deterministic method of genotype phasing and genotype imputation using long haplotypes (Wang et al. 2016). The length of the haplotypes shared between two individuals on a particular chromosome depends on the number of recombinations that occur across the family tree. This number may be known for closely related individuals who share long haplotypes, but it may be unknown for remotely related individuals who share shorter haplotypes. Thus, it is possible to capture family relationships even without pedigree data by searching for long shared haplotypes (Kong et al. 2008). However, the use of pedigree data leads to a more accurate genotype phasing, especially if the chip we impute is of lower density (Daetwyler et al. 2011). The importance of pedigree information decreases with increasing chip density, as more markers increase the recombination resolution and increase the likelihood of finding the correct shared haplotypes (Sargolzaei et al. 2008). The family-based imputation algorithm is iterative and accumulates information by browsing the family tree in each iteration. If both parents are genotyped, and their haplotypes are assembled, imputation occurs directly. The offspring haplotypes are compared to the parent haplotypes, and the missing values are added based on the identified match. If one parent is not genotyped, the nearest genotyped ancestor is searched for based on the pedigree. Genotyped parents have no pedigree as their ancestry genotypes no longer provide the necessary information (Sargolzaei et al. 2014). Haplotypes tend to shorten over generations, mainly due to recombination and mutations. Long shared haplotypes are the result of recent recombinations and mutations and can only be observed among close relatives (Hirschhorn and Daly 2005). The accuracy of haplotype matching between individuals is influenced by their length; the longer the shared haplotype, the more accurate the match (Kong et al. 2008). If more than one haplotype match is found, a higher frequency match is considered the most likely match. Therefore, the search for matching haplotypes using OSW is a suitable method for genotyping and missing marker imputation (Sargolzaei et al. 2014). Using this method, a chromosome or a selected part of the genome is first passed through a large window many times, which gradually diminishes with each new one, while the size of the window is always the same in one run. The maximum window size is set to 1 000 SNPs and the minimum to two SNPs. If pedigree information is available, individuals with a high-density genotype and a phased genotype based on pedigree information can be included in the reference population to improve imputation accuracy. This is a combination of population imputation and pedigree-based imputation methods (Sargolzaei et al. 2014).

Verification of imputation accuracy

Several factors influence imputation accuracy. In addition to the quality of individual genotypes, the size of the reference population of genotyped animals and the number of missing SNPs play a significant role. How these two major factors affect imputation accuracy was described in the study by Kranjcevicova et al. (2019). Where several simulation calculations were performed to monitor the imputation accuracy with varying sizes of the reference population and the number of missing SNPs, the accuracy itself is monitored using the genotype masking method. This method consists of artificially generating the missing SNP value as described by Carvalheiro et al. (2014) and Gurgul et al. (2014). The masked SNPs are then imputed, and the im-

https://doi.org/10.17221/83/2020-CJAS

putation accuracy is calculated as the correlation between the original SNP value before masking and the SNP value after imputation according to the following formula (Kranjcevicova et al. 2019):

$$\text{Imputation accuracy} = \frac{\text{cov}\,(SNP_{BD}, SNP_{AI})}{\sigma_{SNP_{BD}} \times \sigma_{SNP_{AI}}} \quad (1)$$

where:

SNP_{BD} – the single nucleotide polymorphism code before masking;

SNP_{AI} – the single nucleotide polymorphism code after the imputation;

σ – standard deviation.

Another precision parameter is the percentage of imputation success, as a percentage of matches between the original and the imputed SNP (Carvalheiro et al. 2014). In general, studies of imputation accuracy suggest that the reference population size plays a significant role; if it is sufficient, the accuracy is high despite a larger percentage of missing markers (Carvalheiro et al. 2014; Kranjcevicova et al. 2019).

Impact of imputation methods in genomic selection

The method of data imputation was created to compare chips with different density. The main goal of this calculation procedure is to enable all chips to provide a common set of SNP. Berry and Kearney (2011) reported that it is possible to use commercial low-density chips from Illumina and then impute them to a denser Illumina 50K BeadChip. In this case, imputation will facilitate the reduction of genotyping costs and increase the number of genotyped animals in the population. The increased amount of genomic information has also helped to increase the reliability of the prediction of breeding values. Denser chips, such as the BovineHD Genotyping BeadChip, are designed to analyse up to 777 000 SNPs (illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf). The original intention of imputation meant cost savings for genotyping the population. The strategy was to genotype a part of the population on LD chips and then impute them to a reference population genotyped on a higher density chip. This approach was to maximize genetic gain while minimizing the cost. However, the cost of genotyping and the accuracy of imputation are affected by several factors. The accuracy of the imputation itself and the possible loss of information due to incomplete imputation may result in a less accurate estimate of Mendelian sampling but it does not affect the parental average. A study on the strategy of cheap genotyping of pig populations reported a reduction in the cost of genotyping from 120 USD (price of HD chip) to a price in the range of 20.58–34.84 USD per individual if imputation is used (Huang et al. 2012). Imputation is important, especially in studies where it is necessary to increase the size of the population genotyped on HD chips or the sequencing population (Larmer et al. 2017). Information from the 1 000 bull genomes project, which gathers a set of sequenced bulls across cattle breeds, is commonly used to find given haplotypes and sequences (Daetwyler et al. 2014). This project started in 2012 using a database containing 234 sequenced bulls from three breeds of cattle and it has now been enlarged to 3 800 animals and more than 150 breeds of cattle (1000 Bull Genomes Project 2012), mainly due to imputation from SNP chips. The project found that the relatively accurate imputation of sequence genotypes can be achieved, especially for SNPs with a high minor allele frequency or for common allele variants. In SNPs with a low minor allele frequency, i.e., rare alleles, the imputation accuracy was low (Meuwissen et al. 2016). Druet et al. (2013), based on simulation calculations, mentioned the possibility of increasing the accuracy of the imputation of rare alleles in the case of sequencing a large number of ancestors to maximize the number of haplotypes available in the reference population. Another fact influencing the accuracy of imputation in this case is the selection of animals for sequencing into a reference population. Individuals from the reference population must be genetically more closely related to other animals in the population, especially those to be imputed from SNP chips (Stachowicz et al. 2013). Better results of imputation accuracy have also been shown in the case of a two-step approach, where commonly used 50K chips are not directly imputed to the sequenced genotype. They are first imputed to 800K chips, from which they are imputed to sequences in the next phase (van Binsbergen et al. 2014). Higher imputation accuracy in the two-step approach was also observed in the study by VanRaden et al. (2013), who performed an imputation from 3 000 SNPs (3K) to 777 000 SNPs (HD). Fifty thousand

https://doi.org/10.17221/83/2020-CJAS

SNPs were used as an intermediate step. Compared to direct imputation from 3K to HD, the accuracy increased by 2% (VanRaden et al. 2013). According to VanRaden et al. (2013), the incorporation of the HD chip into the genomic evaluation of animals brings on average the only 0.4% increase in reliability using the 50K chip. High density chips provide opportunities in genome-wide association studies. Due to the higher density, it is possible to detect SNPs closer to the quantitative trait locus (QTL) (Howie et al. 2011). A combination of data obtained from classical genotyping and whole-genome sequencing data is increasingly used to find QTLs in relation to production traits. Imputation of these data often helps to identify causative mutations for monogenic traits (Daetwyler et al. 2014). The use of whole-genome sequencing data may lead to higher accuracy in association studies as well as better predictions of genomic breeding values. However, it is necessary to have large data sets with sequenced animal genomes. Imputation from classical SNP chips such as the Illumina Bovine50K BeadChip and Illumina BovineHD BeadChip to whole-genome sequences represents a cheaper data option (van Binsberger et al. 2014). Frichknecht et al. (2017) detected three QTLs for early and late lactation fat content due to data imputation when combining SNP chip information and whole-genomic sequencing information. Genomic data is also widely used in studies of complex health traits. In mastitis, which is one of the costliest diseases in milk production, 22 QTLs were identified using imputation methods. These QTLs explained 14% of the variability of breeding values for resistance to clinical mastitis (Cai et al. 2018).

CONCLUSION

The use of genomic data in the breeding of dairy cattle has helped to increase genetic gain. Genomic data through SNP information has become an integral part of breeding work, and the analysis is continually improving. There is a constant increase in the genotyped animals and new chips that offer new possibilities of disease detection and improvements in verifying the origin of animals. New methods of comparing genomic data such as missing SNP imputation offer new possibilities of making the genotyping technology less expensive. It is not possible to replace quality genotyping with imputation, as it only serves as a complement to the general genotyping of the population, which leads to an increase in the amount of genomic data in animal evaluation and thus also contributes to higher reliability of the prediction of genomic breeding values. The method of data imputation has also become a useful tool in whole-genome association studies, where data from classical SNP chips and whole-genome sequencing are combined. It is thus possible to increase the chances of identifying essential loci that are related to economically significant production traits. Currently, the cost of genotyping has dropped significantly. The use of LD chips is gradually losing its significance. The price of genotyping on 50K chips is comparable to the original price of LD chips. Therefore, imputation becomes very important when adding SNPs to HD chips or sequences. Due to imputation, it is possible to use older genotypes of animals that cannot be genotyped again.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

1000 Bull Genomes Project [Internet]. [Australia]: DairyBio. 2012 - [cited 2020 Mar 29]. Available from: http://www.1000bullgenomes.com.

Aguilar I, Misztal I, Johnson D, Legarra A, Tsuruta S, Lawlor T. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci. 2010 Feb;93(2):743-52.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016 Jan 5;17:81-92.

Bauer J, Pribyl J, Vostry L. Contribution of domestic and Interbull records to reliabilities of single-step genomic breeding values in dairy cattle. Czech J Anim Sci. 2015 Mar;60(3):263-7.

Berry DP, Kearney JF. Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. Animal. 2011 Jun;5(8):1162-9.

Browning B, Browning S. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009 Feb 13;84(2):210-23.

Browning B, Browning S. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics. 2013 Jun 1;194(2):459-71.

https://doi.org/10.17221/83/2020-CJAS

Cai Z, Guldbrandtsen B, Lund MS, Sahana G. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. BMC Genomics. 2018 Sep 6;19(1): [11 p.].

Carvalheiro R, Solomon B, Neves HR, Sargolzaei M, Schenkel FS, Utsunomiya YT, O'Brien AM, Solkner J, McEwan JC, Van Tassel CP, Sonstegard TS, Garcia JF. Accuracy of genotype imputation in Nelore cattle. Genet Sel Evol. 2014 Oct 10;46: [11 p.].

Cheung CY, Thompson EA, Wijsman EM. GIGI: An approach to effective imputation of dense genotypes on large pedigrees. Am J Hum Genet. 2013 Apr 4;92(4):504-16.

Christensen O, Lund M. Genomic prediction when some animals are not genotyped. Gen Sel Evol. 2010 Jan 27; 42(1): [8 p.].

Cooper TA, Wiggans GR, VanRaden PM. Short communication: Relationship of call rate and accuracy of single nucleotide polymorphism genotypes in dairy cattle. J Dairy Sci. 2013 May;96(5):3336-9.

Daetwyler H, Wiggans G, Hayes B, Woolliams J, Goddard M. Imputation of missing genotypes from sparse to high density using long-range phasing. Genetics. 2011 Sep 1;189(1):317-27.

Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, Liao X, Djari A, Rodriguez SC, Grohs C, Esquerre D, Bouchez O, Rossignol MN, Klopp C, Rocha D, Fritz S, Eggen A, Bowman PJ, Coote D, Chamberlain AJ, Anderson C, VanTassell CP, Hulsegge I, Goddard ME, Guldbrandtsen B, Lund MS, Veerkamp RF, Boichard DA, Fries R, Hayes BJ. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014 Jul 13;46(8):858-65.

Druet T, Georges M. A hidden Markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. Genetics. 2010 Mar 1;184(3):789-98.

Druet T, Macleod IM, Hayes B. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of prediction. Heredity. 2013 Apr 3;112(1):39-47.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawaniti K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencinc (GBS) approach for high diversity species. PloS One. 2011 May 4;6(5): [10 p.].

Frischknecht M, Pausch H, Bapst B, Signer-Hasler H, Flury C, Garrick D, Stricker C, Fries R, Gredler-Grandl B. Highly accurate sequence imputation enables precise QTL mapping in Brown Swiss cattle. BMC Genomics. 2017 Dec 29;18(1): [10 p.].

Garcia-Ruiz A, Cole J, VanRaden P, Wiggans G, Ruiz-Lopez F, Van Tassell C. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. PNAS. 2016 Jul 12;113(28):E3995-4004.

Gurgul A, Sienko K, Zukowski K, Pawlina K, Bugno-Poniewierska M. Imputation accuracy of bovine spongiform encephalopathy-associated PRNP indel polymorphisms from middle-density SNPs arrays. Czech J Anim Sci. 2014 May;59(5):224-9.

Hayes B, Chamberlain A, Goddard M. Use of markers in linkage disequilibrium with QTL in breeding programs. In: Proceedings of the 8th World Congress on Genetics Applied to Livestock Production; 2006 Aug 13-18; Belo Horizonte, Minas Gerais, Brazil: Instituto Prociencia; 2006. p. 30-6.

Hayes B, Bowman P, Chamberlain A, Goddard M. Invited review: Genomic selection in dairy cattle. J Dairy Sci. 2009a Feb;92(2):433-43.

Hayes BJ, Visscher PM, Goddard ME. Increase accuracy of artificial selection by using the realized relationship matrix. Genet Res. 2009b Feb 17;91(1):47-60.

Hickey J, Kinghorn BP, Tier B, van der Werf JHJ, Cleveland MA. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. Gen Sel Evol. 2012 Jun;44(9):1-11.

Hirschhorn J, Daly M. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005 Feb 1;6(2):95-108.

Howie BN, Donnelly P, Marchini J, Schork NJ. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009 Jun 19;5(6): [15 p.].

Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3. 2011 Nov 1;1(6):457-70.

Huang Y, Hickey JM, Cleveland MA, Maltecca C. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. Gen Sel Evol. 2012 July 31;44(1): [8 p.].

Jenko J, Wiggans GR, Cooper TA, Eaglen SAE, Luff WGL, Bichard M, Pong-Wong R, Woolliams JA. Cow genotyping strategies for genomic selection in a small dairy cattle population. J Dairy Sci. 2017 Jan;100(1):439-52.

Khatkar MS, Zenger KR, Hobbs M, Hawken RJ, Cavanagh JAL, Barris W, McClintock AE, McClintock S, Thomson PC, Tier B, Nicholas FW, Raadsma HW. A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in Holstein-Friesian cattle. Genetics. 2007 Jun 1; 176(2):763-72.

Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI, Ingason A, Steinberg S, Rafnar T, Sulem P, Mouy M, Jonsson F, Thorsteinsdottir U, Gudbjartsson DF, Stefansson H, Stefansson K. Detection

https://doi.org/10.17221/83/2020-CJAS

of sharing by descent, long-range phasing and haplotype imputation. Nat Genet. 2008 Aug 17;40(9):1068-75.

Kranjcevicova A, Kasna E, Brzakova M, Pribyl J, Vostry L. Impact of reference population and marker density on accuracy of population imputation. Czech J Anim Sci. 2019 Oct;64(10):405-10.

Larmer A, Sargolzaei M, Brito LF, Ventura RV, Schenkel FS. Novel methods for genotype imputation to whole-genome sequence and simple linear model to predict imputation accuracy. BMC Genet. 2017 Dec 27;18(1): [12 p.].

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010 Dec;34 (8):816-34.

Meuwissen T, Hayes BJ, Goddard ME. Genomic selection: A paradigm shift in animal breeding. Anim Front. 2016 Jan 1;6(1):6-14.

Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci. 2009 Sep; 92(9):4648-55.

Misztal I, Aguilar I, Legarra A, Lawlor TJ. Choice of parameters for single-step genomic evaluation for type. In: ASAS Annual Meeting; 2010 Jul 11-15; Tolouse, France. [Tolouse, France]: ASAS; 2010. p. 533.

Misztal I, Aggrey S, Muir W. Experiences with a single-step genome evaluation. Poult. 2013 Sep 1;92(9):2530-4.

Nicolazzi EL, Biffani S, Jansen G. Short communication: Imputing genotypes using PedImpute fast algorithm combining pedigree and population information. J Dairy Sci. 2013 Apr;96(4):2649-53.

Nicolazzi E, Picciolini M, Strozzi F, Schnabel R, Lawley C, Pirani A, Brew F, Stella A. SNPchiMp: A database to disentangle the SNPchip jungle in bovine livestock. BMC Genomics. 2014 Feb 11;15(1): [6 p.].

Robledo D, Palaiokostas C, Bargelloni L, Martinez P, Houston R. Applications of genotyping by sequencing in aquaculture breedin and genetics. Rev Aquac. 2017 Feb;10:670-82.

Sargolzaei M, Schenkel F, Jansen G, Schaeffer L. Extent of linkage disequilibrium in Holstein cattle in North America. J Dairy Sci. 2008 May;91(5):2106-17.

Sargolzaei M, Chesnais J, Schenkel F. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014 Jun 17;15(1): [12 p.].

Schaeffer L. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet. 2006 Aug;123(4): 218-23.

Stachowicz K, Larmer S, Jamrozik J, Moore SS, Miller SP. Sequencing and genotypng for the whole genome selection in Canadian beef populations. Proc Assoc Advmt Anim Breed Genet. 2013 Oct;20:344-7.

Starkey M, Elaswarapu R. Genomics: Essential methods. Hoboken: Wiley; 2010. 360 p.

van Binsbergen R, Bink MCAM, Calus MPL, van Eeuwijk FA, Hayes BJ, Hulsegge I, Veerkamp RF. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. Gen Sel Evol. 2014 Jul 15;46(1): [13 p.].

VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008 Nov;91(11):4414-23.

VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J, Schenkel F. Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci. 2009 Jan;92(1):16-24.

VanRaden P, O'Connell J, Wiggans G, Weigel K. Genomic evaluations with many more genotypes. Gen Sel Evol. 2011 Mar 2;43(1): [11 p.].

VanRaden PM, Null DJ, Sargolzaei M, Wiggans GR, Tooker ME, Cole JB, Sonstegard TS, Connor EE, Winters M, van Kaam JBCHM, Valentini A, Van Doormaal BJ, Faust MA, Doak GA. Genomic imputation and evaluation using high-density Holstein genotypes. J Dairy Sci. 2013 Jan; 96(1):668-78.

Walker EJ, Siminovitch KA. Primer: Genomic and proteomic tools for the molecular dissection of disease. Nat Clin Pract Rheumatol. 2007 Oct;3:580-9.

Wang Y, Lin G, Li C, Stothard P. Genotype imputation methods and their effects on genomic predictions in cattle. Springer Sci Rev. 2016 Feb;4:79-98.

Wang X, Su G, Hao D, Lund MS, Kadarmideen HN. Comparisons of improved genomic predictions generated by different imputation methods for genotyping by sequencing data in livestock populations. J Anim Sci Biotechnol. 2020 Jan 7;11(1): [12 p.].

Weigel K. Genomic selection of dairy cattle: A review of methods, strategies, and impact. J Anim Breed Genet. 2017 Dec;1(1):1-15.

Wiggans G, Cooper T, VanRaden P, Van Tassell C, Bickhart D, Sonstegard T. Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle. J Dairy Sci. 2016 Jun;99(6):4504-11.

Wiggans G, Cole J, Hubbard S, Sonstegard T. Genomic selection in dairy cattle: The USDA experience. Annu Rev Anim Biosci. 2017 Feb;5(1):309-27.

Zhang Z, Druet T. Marker imputation with low-density marker panels in Dutch Holstein cattle. J Dairy Sci. 2010 Nov;93(11):5487-94.

Received: April 5, 2020
Accepted: December 10, 2020