Chapter 16

Genomic Prediction of Complex Traits in Animal Breeding with Long Breeding History, the Dairy Cattle Case

Joel Ira Weller

Abstract

In accordance with the infinitesimal model for quantitative traits, a very large number of genes affect nearly all economic traits. In only two cases has the causative polymorphism been determined for genes affecting economic traits in dairy cattle. Most current methods for genomic evaluation are based on the "two-step" method. Genetic evaluations are computed by the individual animal model, and functions of the evaluations of progeny-tested sires are the dependent variable for estimation of marker effects. With the adoption of genomic evaluation in 2008, annual rates of genetic gain in the US increased from ~50–100% for yield traits and from threefold to fourfold for lowly heritable traits, including female fertility, herd-life and somatic cell concentration. Gradual elimination of the progeny test scheme has led to a reduction in the number of sires with daughter records and less genetic ties between years. As genotyping costs decrease, the number of cows genotyped will continue to increase, and these records will become the basic data used to compute genomic evaluations, most likely via application of "single-step" methodologies. Less emphasis in selection goals will be placed on milk production traits, and more on health, reproduction, and efficiency traits and "environmentally friendly" production. Genetic variance for economic traits is maintained by increase in frequency of rare alleles, new mutations, and changes in selection goals and management.

Key words Genomic prediction, Genomic selection, Dairy cattle, Animal breeding, Complex traits

1 Introduction

Nearly all traits of economic importance in dairy cattle are recorded only on females. All breeding programs must deal with the anomaly that male fertility is nearly unlimited, while cows, without biotechnological intervention, produce at best a single calf per year. Beginning in the 1950s "progeny test" schemes were implemented in dairy cattle breeding programs in most developed countries. An example for the Israeli dairy cattle population, consisting of ~120,000 cows, nearly all of the Holstein breed, is diagramed in Fig. 1 [1]. Prior to genomic selection, a cohort of ~50 young bulls were each mated to a sample of 500–1000 cows in order to produce

Nourollah Ahmadi and Jérôme Bartholomé (eds.), *Genomic Prediction of Complex Traits: Methods and Protocols*, Methods in Molecular Biology, vol. 2467, https://doi.org/10.1007/978-1-0716-2205-6_16,

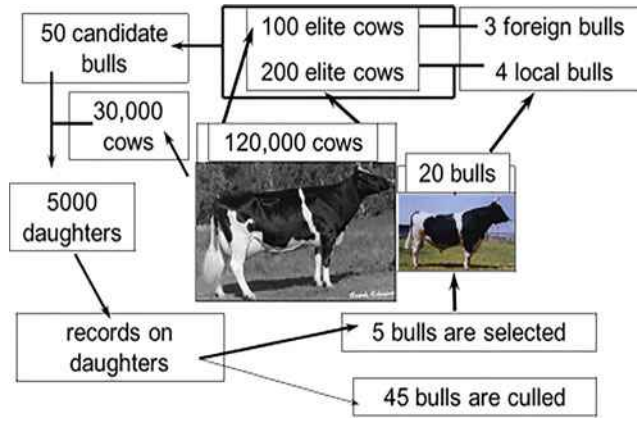

Fig. 1 The Israeli dairy cattle breeding program prior to the introduction of genomics. Each year ~5 young bulls are selected for general service, and are used on the average for 4 years. Thus, ~20 bulls are in general service each year. Only ~4 of these bulls are used as bull sires, and mated to 200 elite cows. The remaining elite cows are mated using imported semen of foreign bulls. Elite cows and bulls for general service are selected based on their genetic evaluations for the Israeli breeding index. The genetic evaluations of the young bulls are based chiefly on the records of their daughters, ~100 daughters/bull. Genetic evaluations are computed three times yearly

50–100 daughters per young bull. Approximately 10% of the bulls with the highest genetic evaluations, based on the performance of their daughters, were then selected for mating to the general population. The main disadvantage of the progeny test scheme is that, although bulls reach sexual maturity at the age of 1 year, bulls are approximately 5 years old by the time that progeny test records from the first crop of daughters is available. In addition, the number of bulls that can be progeny tested is limited, and will generally be dependent on the total size of the milk-recorded population [1]. Annual genetic gain is computed as the sum of genetic gain per generation over the four paths of inheritance; sire-to-son, sire-to-daughter, dam-to-son, and dam-to-daughter; divided by the sum of the four generation intervals.

Studies proposing the application of marker-assisted selection in dairy cattle were first published in the 1980s (reviewed by [1]). Nearly all studies assumed that the main gain of marker-assisted selection would be extensive use of young bulls based on both pedigree and marker information. Thus, the mean generation interval from sires to their progeny would be reduced, increasing the mean genetic gain per year. This scheme only becomes practical if the reliabilities of genomic evaluations of young bulls are significantly higher than their reliabilities based only on pedigree [1]. ("Accuracy" is defined as the correlation between the genetic evaluation and the actual breeding value, and "reliability" is defined

as the square of the accuracy, or the coefficient of determination of the evaluation relative to the actual breeding value [1].) The early studies assumed that a relatively small number of chromosomal segments would be followed by <100 markers, e.g., Hayes and Goddard [2]. Several breeding programs based on this strategy were in fact applied to dairy cattle beginning in the first decade of this century [3, 4].

With the introduction of high-density SNP chips in 2008 including >50,000 markers [5], genomic selection became a reality. By 2014, this technology was successfully implemented in the United States, Canada, Great Britain, Ireland, New Zealand, Australia, France, the Netherlands, Germany, and the Scandinavian countries [6]. Adoption of genomic selection breeding programs in the major dairy producing countries has led to significant changes in the worldwide dairy industry.

In this chapter, I will first describe the methods used for genetic evaluation of dairy cattle prior to the introduction of genomics. I will then discuss what we now know about the nature of genetic variation for economic traits, methods to compute genomic evaluations based on high-density SNP chips, factors that affect the accuracy of genomic evaluations, the changes that have occurred in commercial breeding programs due to the introduction of genomic evaluations, evaluation of actual genetic progress as compared to expectations, and make predictions of future directions.

2 Genetic Evaluation of Dairy Cattle Prior to Genomics

Unlike model species and most agricultural plants, genetic evaluation of dairy cattle has always been based on analysis of field records from commercial herds with different management systems and climate. Until the 1980s, routine genetic evaluations for nearly all major dairy cattle breeding programs were based on "sire models" [1]. That is the dependent variable in the prediction model was the cow's first parity production, and the independent variable of interest was the sires' genetic value. Since the distributions of the sires' daughters across herds was not orthogonal, it was necessary to correct for the herd, or herd-year-season (HYS) effect. Usually this effect was absorbed to reduce the number of equations. Considering the computing limitations prior to this century, and that large dairy populations included thousands of sires, these systems of equations could only be solved by iteration. Sire models did not account for relationships other than between the cow and her sire, and did not provide genetic evaluations for the individual cows [1].

In 1976, Henderson devised a simple algorithm to compute the inverse of the numerator relationship matrix from a list of animals and their parents [7]. Based on this algorithm, and the increase in computing capabilities, it became possible to apply

"animal models" [1] for routine genetic analysis. A simple individual animal model (IAM) is given in Eq. 1:

$$Y_{ijk} = H_i + a_j + p_j + e_{ijk} \quad (1)$$

where Y_{ijk} is record k of individual j in HYS i, H_i is the fixed effect of HYS i, a_j is the random additive genetic effect of individual j, p_j is the random permanent environmental effect for individual j, and e_{ijk} is the random residual associated with each record.

As noted previously, sire models generally considered only the cows' first parity records, because two or more records on the same cow will have a positive covariance. In the IAM this covariance is taken into account by inclusion of a "permanent environmental" effect, common to all records of each cow, in addition to the additive genetic effect of the cow. Thus, the IAM includes two equations for each cow included in the analysis. Since large dairy populations include millions of cows, the IAM was not computationally feasible until the 1980s. In a fixed model, the additive genetic and permanent environmental effects would be completely confounded, because each level of these two effects refer to the same individual. In the IAM these effects can be estimated separately, because both are assumed to be random, and their variance structures are different. The variance matrix for the permanent environmental effect will be $\boldsymbol{I}\sigma_p^2$, where $\boldsymbol{I}$ is an identity matrix and σ_p^2 is the variance component of the permanent environmental effect. The variance matrix for the additive genetic effect will be $\boldsymbol{A}\sigma_a^2$, where $\boldsymbol{A}$ is the numerator relationship matrix and σ_a^2 is additive genetic variance. (Throughout this chapter, I will employ the conventions that matrices will be denoted by upper case letters, vectors by lower case letters, and both will be denoted in *bold type*. The transpose of a matrix will be denoted by an apostrophe and the inverse of a matrix by the minus one power.) For most economic traits in dairy cattle, only females will have records. Individuals without records, such as sires of cows or dams of cows without records, will be included in the analysis via the relationship matrix. Thus, in the IAM sire evaluations are derived via the relationship matrix. Even with current computing capabilities, a typical commercial IAM containing millions of equations can only be solved by iteration [1].

The IAM as described would accurately reflect reality if the pedigree of all animals could be traced back to a group of unrelated animals [1]. However, this is never the case. Various animals of different ages will have missing pedigree information. Thus, a fixed "genetic group" effect is usually included in the model to account for genetic trend and other genetic effects not included in the known genetic relationships. Thompson [8] proposed a grouping strategy based on "phantom parents." Each individual with unknown parents is assigned to phantom parents. These phantom

parents are then assigned into groups based on year of birth, sex, and whether the sire, dam or both parents are unknown. The genetic evaluations are then computed as the sum of the additive genetic effects and the group effects of each individual. Westell et al. [9] developed a simple algorithm to directly compute estimated breeding values that incorporate the group effects for each individual.

Including the genetic group effect, the IAM of Eq. 1 can now be described in matrix format as follows:

$$y = Hh + Za + ZA_g g + P_p + e \tag{2}$$

where $\boldsymbol{y}$ represents the vector of animal record; $\boldsymbol{h}$, $\boldsymbol{a}$, $\boldsymbol{g}$, and $\boldsymbol{p}$ are vectors of effects for HYS, additive genetic effect, unknown-parent group, and permanent environment, respectively; $\boldsymbol{H}$, $\boldsymbol{Z}$, $\boldsymbol{ZA_g}$ and $\boldsymbol{P}$, are incidence matrices for these effects; and $\boldsymbol{e}$ is the residual variance. The matrix $\boldsymbol{A_g}$ relates animals to unknown-ancestor groups, via the algorithm of Westell et al. [9].

VanRaden and Wiggans [10] demonstrated that important elements of the IAM could be expressed as relatively simple algebraic formulae. The cow's own information is summarized by her "yield deviation" (YD), a weighted average of yields adjusted for effects other than genetic merit and error. Defining $\widehat{\boldsymbol{h}}$ and $\widehat{\boldsymbol{p}}$ as the vectors of solutions for the HYS and permanent environmental effect, each cow's YD is computed as the element of $\boldsymbol{Z}'(\boldsymbol{y}-\boldsymbol{H}\widehat{\boldsymbol{h}}-\boldsymbol{P}\widehat{\boldsymbol{p}})$ for that cow divided by the corresponding diagonal element of $\boldsymbol{Z}'\boldsymbol{Z}$; i.e., a weighted average of the cow's yields adjusted for effects other than genetic merit and residual.

For each bull with daughter records in the analysis, the bull's daughter yield deviation (DYD) is computed by summing over all daughters with records as follows:

$$\mathrm{DYD} = \frac{\sum_{1}^{N} \left(q_{\mathrm{prog}}\ W2_{\mathrm{prog}} \left(\mathrm{YD}_{\mathrm{prog}} - \mathrm{PTA}_{\mathrm{mate}} \right) \right)}{\sum_{1}^{N} q_{\mathrm{prog}}\ W2_{\mathrm{prog}}} \tag{3}$$

where N is the number of daughters of each sire. For each daughter with records, q_{prog} equals 1 if progeny's other parent is known and two-third if unknown. $W2_{\mathrm{prog}}$ = the number of lactations of the cow divided by the sum of the number of lactations of the cow and $2q_{\mathrm{prog}}\left(\sigma_e^2/\sigma_a^2\right)$, where σ_e^2 and σ_a^2 are the residual and additive genetic variances. YD_{prog} is the daughter's YD. $\mathrm{PTA}_{\mathrm{mate}}$ is the predicted transmitting ability of the cow's dam, the bull's mate. PTA is half of the animal's breeding value. Thus, a DYD is the weighted mean of the bull's daughters' YD corrected for the genetic merit of the daughters' dams. This definition of DYD is somewhat simplified, as compared to the definition of VanRaden and Wiggans [10], based on the assumption that all lactation records are weighted equally.

VanRaden and Wiggans [10] wrote: "The DYD may be helpful in explaining evaluations and also as a dependent variable in statistical tests and calculation of conversions across countries." Unlike a simple mean of the cows' records or a simple mean of bull's daughter records, YD and DYD are corrected for the other effects included in the model. However, unlike genetic evaluations which are based on random effects, both YD and DYD are not regressed toward the mean as a function of the additive genetic variance.

3 Detection of Segregating Quantitative Trait Loci (QTL) in Dairy Cattle, the "Granddaughter Design"

With the discovery of DNA level genetic markers in the 1980s and especially microsatellites in 1989 [11] detection of the individual quantitative trait loci (QTL) affecting quantitative traits became possible for all agricultural species of interest. Unlike most plant species, production of experimental dairy cattle populations specifically for QTL detection was not a viable option. As noted previously, prior to genomics, genetic evaluation of dairy cattle was based on the progeny test scheme. Thus, only a small fraction of males is used as parents of the next generation, but each sire has many daughters. Furthermore, these bulls are generally progeny of an even smaller number of grandsires. Based on this population structure, Weller et al. [12] proposed the granddaughter design, which could be applied to large commercial dairy cattle populations. This design is diagrammed in Fig. 2. They proposed genotyping only the grandsires and their progeny-tested sons. Since the genetic evaluation of each son is based on 50–100 daughters, the reliability of their evaluations are close to 90% for a moderately heritable trait, such as milk production. In each grandsire family, for each marker, the sons are divided into two groups based on which grand-paternal allele was passed, and significance of the contrast is tested. This design is able to randomize all sources of variation not linked to a specific chromosomal region [12].

In the first preliminary granddaughter design performed on the Israeli Holstein population, the genetic evaluations of the sires were the dependent variables [13]. Georges et al. [14] first proposed that DYD should be used as the dependent variable for the granddaughter design. The vast majority of published studies based on analysis of bull genotypes have used either DYD or "de-regressed" genetic evaluations as the dependent variable. Formula to compute de-regressed genetic evaluations are given in VanRaden et al. [15]. Analysis of DYD as the dependent variable has the advantage, compared to genetic evaluations, that the variance of DYD decreases with the number of actual records included in the bull's DYD. Thus weighting DYD by the reliabilities of the genetic

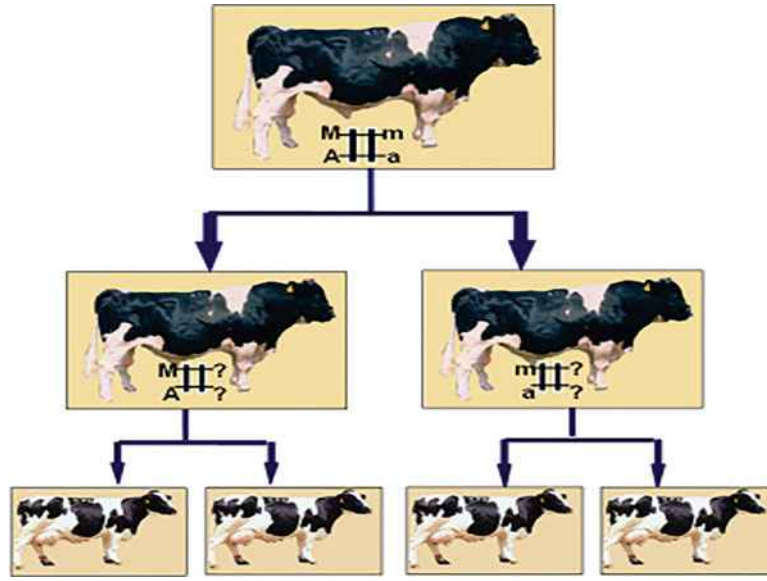

Fig. 2 The granddaughter design to detect segregating quantitative trait loci. A single chromosome is shown, although in practice all chromosomes with genetic markers are analyzed

evaluations should yield "reasonable" results. That is, the weighting factor of a record should be inversely related to the residual variance. In the statistical analysis of the granddaughter design, more weight would be given to DYD based on more daughter records, and these DYD would have lower variances. During the 1990s the granddaughter design was applied to most of the major dairy cattle populations (reviewed by [1]). All of these studies claimed that segregating QTL were detected, and became the basis for the first commercial marker-assisted selection programs [1].

4 The Architecture of Quantitative Genetic Variation, Major Genes vs The Infinitesimal Model

As noted previously, prior to the development of high density SNP-chips, all studies that proposed methods for marker-assisted selection assumed that a relatively low number of chromosomal segments would be followed, each containing a single gene of interest (e. g [16]). It is now clear that the infinitesimal model, first proposed by Fisher [17] seems to be much closer to reality. That is, genetic variation in quantitative traits is determined by a very large number of loci, each with very small effects.

We will briefly review the evidence both from dairy cattle and humans. By 2008, tens of thousands of humans were genotyped for

high density SNP-chips. This data was used to detect genes affecting human height, an easy to measure trait with a heritability of ~90%. The surprising result was that no major genes were detected, and that those genes that were detected only accounted for a small fraction of the total additive genetic variance. Maher [18] discussed this anomaly in detail and proposed various explanations. Visscher [19] was able to demonstrate that most of the genetic variance could be explained by analysis of chromosomal segments, rather than specific genes. That is most of the additive genetic variance was due to a very large number of genes with effects too small to be detected even with a sample of tens of thousands of individuals.

The second line of evidence comes from the results of SNP chips in dairy cattle. The first practical methods for genomic evaluation were based on estimation of the effects of all valid markers on the trait of interest, and summation of these effects to obtain the genomic evaluation for each individual [20]. Assuming a relatively low number of genes affecting any specific trait, the estimated effects associated with most markers should only represent random variation. Thus, it should be possible to select a subset of markers that only include those markers actually in linkage disequilibrium with genes affecting the trait analyzed. Determination of genomic evaluations by this subset of markers should be more accurate, due to the reduction of "random noise" generated by markers with no effects. In practice, genomic evaluation based on selected subsets of markers has at best been only marginally better than evaluation on all markers, and in most cases not nearly as efficient [21].

The third line of evidence comes from the causative polymorphisms, the "quantitative trait nucleotides" (QTN) that have been determined in dairy cattle. The search for QTLs has been in high gear for nearly 20 years. Yet, only 2 QTN have been found that meet all the criteria for positive detection of a QTN proposed by Ron and Weller [22], and Weller and Ron [23]; DGAT1 that chiefly affects milk fat concentration [24]; and ABCG2, which chiefly affects protein concentration [25].

The current release of the cattle QTL data base contains 160,659 cattle QTLs curated from 1030 publications and represent 675 different cattle traits [26] (https://www.animalgenome.org/cgi-bin/QTLdb/index). Due to differing trait definitions, possible genetic differences between populations and the relatively large confidence intervals for QTL location, it is unclear as to how many of these associations can be considered confirmed by independent studies. In the modified granddaughter design of Weller et al. [27], genotypes for 83 grandsires and 17,217 sons were determined by imputation for 3,148,506 polymorphisms across the entire genome. A total of 30 trait-by-chromosomal segment effects segregating in the US Holstein population with probabilities of $<10^{-20}$ to accept the null hypotheses of no segregating gene affecting the trait within the chromosomal segment were found.

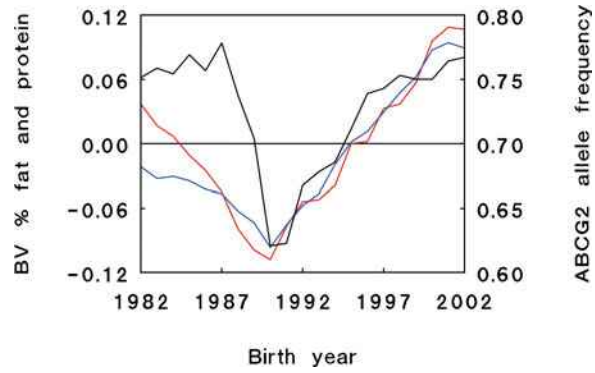

Fig. 3 Genetic trends for protein and fat concentration and frequency of the *ABCG2* 581Y allele in the Israeli Holstein cow population by birth year. —, *ABCG2* 581Y allele frequency; —, mean yearly breeding values for fat percentage; —, mean yearly breeding values for protein percentage

However, they were not able to conclusively determine the QTN for any additional loci, and none of the effects found explained more than a few percent of the genetic variance.

DGAT1 and ABCG2 can be considered the exceptions that prove the rule. Both fat and protein concentration have heritability greater than 50%, and are traits that have not been under direct selection. The polymorphisms for DGAT1 has been maintained in the population, because the effect of allele substitution is approximately neutral with respect to most selection indices; the allele that increases fat concentration also decreased protein production [22]. With respect to ABCG2 the allele that increases protein concentration reduces fluid milk production [22]. Until 1980, selection in most countries was chiefly for fluid milk, which resulted in an increase in the allele that increases milk volume. Since then protein production has become the main trait for selection in most countries, and the allele that increases protein concentration has reached fixation in most commercial populations [28]. Genetic trends for protein and fat concentration and frequency of the *ABCG2* 581Y allele, the allele that causes increased protein and fat concentration, in the Israeli Holstein cow population are plotted by birth year in Fig. 3.

5 Computation of Genomic Evaluations in Dairy Cattle Based on Analysis of Sire Evaluations

The mathematically optimal solution for computation of genomic evaluations for dairy cattle based on the IAM would be to include a marker effect for each SNP into the animal model. The genomic evaluation for each animal would then be the sum of the genetic groups and additive genetic effect given in Eq. 2, which would

include all genetic effects not linked to the markers, and the sum of the effects of all genotyped markers. This model cannot be applied in practice, because only a small fraction of the animals included in the standard evaluations of commercial dairy populations have genotypes; and in the early years, these were chiefly sires, who are included the IAM analysis only through the relationship matrix.

The first practical method for genomic evaluation in dairy cattle was proposed by VanRaden [20], using DYD as the dependent variable. The vector of sires' DYD, $\boldsymbol{y}$, can be modeled as follows:

$$y = xb + Zu + e \tag{4}$$

where $\boldsymbol{b}$ is the mean (the only fixed effect in the model); $\boldsymbol{x}$ is a vector with all elements equal to one; $\boldsymbol{u}$ are the marker effects, which are assumed to be random; $\boldsymbol{Z}$ is the matrix that relates marker effects to the individual records; and $\boldsymbol{e}$ is the random error vector with variance matrix $\boldsymbol{R}$. All markers were assumed to be biallelic.

The positive covariance among markers in genetic linkage on the same chromosome is taken into account via the $\boldsymbol{Z}$ matrix, which is computed as follows. Let $\boldsymbol{M}$ be the matrix that specifies which marker alleles each individual inherited. Dimensions of $\boldsymbol{M}$ are the number of individuals by the number of markers. If elements of $\boldsymbol{M}$ are set to -1, 0, and 1 for the homozygote, heterozygote, and other homozygote, respectively; diagonals of $\boldsymbol{MM}'$ count the number of homozygous loci for each individual, and off-diagonals measure the number of alleles shared by relatives. Let the frequency of the second allele at locus i be p_i, and let the matrix $\boldsymbol{P}$ contains allele frequencies expressed as a difference from 0.5 and multiplied by 2, so that column i of $\boldsymbol{P}$ is $2(p_i - 0.5)$. $\boldsymbol{Z}$ is then defined as $\boldsymbol{M} - \boldsymbol{P}$, so that mean values of the allele effects is $\boldsymbol{Z} = 0$. VanRaden [20] assumed that allelic frequencies would be computed from the "base animals." That is animals with genotypes, but without ancestors with genotypes. Aguilar et al. [29] investigated this question for "single-step" models and concluded that assuming $p_i = 0.5$ resulted in optimal genomic evaluations.

In single trait analysis by the mixed model, it is generally assumed that the residual matrix is equal to the identity matrix times a constant. This will not be the case for DYD, since DYD of related bulls will have a positive covariance. Various studies have therefore assumed a variance matrix equal to the relationship matrix among bulls multiplied by a constant. In addition, the residuals will be a function of the number of daughter records per bull. The standard procedure to account for this is to weight the DYD by the reliabilities of the evaluations. Although this may be approximately correct, over the general range of sire reliabilities, it is not clear what should be done with bulls with thousands of daughters. These bulls should have residual variances approaching zero, and it is unknown how the generalized linear model should behave in this

situation. In the model of VanRaden [20] diagonals of $\boldsymbol{R}$ were computed as: $(1/R_{\text{dau}} - 1)\sigma_e^2$, where R_{dau} is the bull's reliability from daughters with parent information excluded, and $\sigma^2{}_e$ is the residual variance of the DYD not explained by the marker effects. Off-diagonals were assumed to be zero, based on the assumption that most of the common genetic effects among relatives would be included in the marker effects.

VanRaden [20] gives three different methods to derive genomic breeding values, but only one will be presented here. In this method the vector of the effects of the individual markers, $\boldsymbol{u}$, can be derived from the following equations:

$$\left[\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{I}\lambda\right][\boldsymbol{u}] = \boldsymbol{Z}'\boldsymbol{R}^{-1}\left(\boldsymbol{y} - \boldsymbol{x}\widehat{\boldsymbol{b}}\right) \tag{5}$$

where $\lambda =$ the ratio σ_e^2/σ_u^2, which equals the sum across marker loci $2\ \Sigma\ p_i(1 - p_i)$ times the ratio σ_e^2/σ_a^2, where σ_e^2, σ_u^2 and σ_a^2 are the residual, marker and total additive genetic variance, and $\widehat{\boldsymbol{xb}}$ are the solutions for the means of $\boldsymbol{y}$. Genomic breeding values are then obtained as $\boldsymbol{Z}\widehat{\boldsymbol{u}}$, where $\widehat{\boldsymbol{u}}$ are the solutions for the marker effects. As noted previously, all off-diagonal elements of $\boldsymbol{R}$ are assumed to be zero. Therefore, $\boldsymbol{R}^{-1}$ can be computed by inverting each diagonal element. Thus, this method is computationally tractable, as it is not necessary to invert any large matrices, and Eq. 5 can be solved by standard iteration algorithms, such as Gauss-Seidel [20].

Final genomic estimate of breeding values (GEBV) are derived by selection index which includes three factors: (1) the estimated direct genomic evaluations as described, (2) the parent average genetic evaluations computed from the subset of genotyped ancestors using known relationships, and (3) parent average genetic evaluations for individuals with genetic evaluations for both parents, or pedigree indices of individuals without dam genetic evaluations.

DYD are only computed for traits analyzed by the single trait IAM. Thus DYD are not computed for fat and protein concentration, for which genetic evaluations are usually derived from the genetic evaluations of milk, fat, and protein yield. DYD are also not computed for alternative analysis models, such as a multi-trait IAM in which each parity is considered as a separate trait [30]. In these cases, VanRaden et al. [15] suggested analysis of de-regressed genetic evaluations as the dependent variable.

6 "Single-step" Methods of Genomic Evaluation

As noted previously, until recently, only a small fraction of cows with records were genotyped. This situation is quickly changing. By March of 2021, more than 3.5 million North American Holstein cows were genotyped (https://queries.uscdcb.com/Genotype/

cur_density.html). Methods to compute genomic evaluations based on incorporation of marker effects into the IAM, accounting for the fact that only a small fraction of the animals were genotyped were developed by Legarra et al. [31]. This method was termed "single step," as opposed to the method of VanRaden [20], which was termed "two step." Mathematical description of the single-step model accounting for the fact that only a small fraction of animals has genotypes is quite complex and beyond the scope of this chapter. For extensive description of the single-step model *see*: refs. 29, 31–33. I will summarize the advantages and disadvantages of both methods, based on theoretical considerations, and results on analysis of actual data.

Advantages of the two-step method include no change to the regular evaluations and simple steps for predicting genomic values for young genotyped animals. Furthermore, overall computing time will be considerable less than for single-step methods. The main disadvantage is that genotypes of cows are not included in the analysis. Models have been proposed that included both bull and cow genotypes with daughter yields as the dependent variable for cows, but their application to actual data have not resulted in more accurate evaluations [34]. Additional disadvantages include weighting parameters, such as variance components [35] or selection index coefficients [15], loss of information, and biased evaluations [29, 32]. Furthermore, the extension to alternative analysis models, such as multi-trait evaluations or test-day models is not obvious. As noted previously, there are several problems in the use of DYD as the dependent variable. These problems are weights (caused by different amount of information in the original data set), bias (caused by selection, for example), accuracy (for animals in small herds), and collinearity (for example, the yield deviations of two cows in the same herd). Furthermore, if genomic selection is used, the expectation of Mendelian sampling in selected animals is not zero, which can lead to biased evaluations [36]. Although single-step methodologies appear to be superior on theoretical considerations, and can be readily applied to any analysis model, differences in accuracy of prediction between the two methods on actual data so far are at best minimal [35]. In addition, single-step methods have three drawbacks: (a) They require much more extensive computing, due to the generally huge number of equations included in the analysis model. (b) It is necessary to partition the genetic variance between the fraction associated with markers and the remainder that is independent of the marker effects. Although this factor can theoretically be computed for historical data, it is unknown for animals that have not yet produced records. (c) Convergence to solutions by iteration is problematic [33].

7 The Factors that Affect the Accuracy of Genomic Evaluations

Nearly all studies that compared genomic evaluations to traditional genetic evaluations based only on phenotypic records and relationships have assumed that the basis for comparison for young animals without records or progeny records is the animal's parent average of genetic evaluations (PA) derived by standard mixed model methodology. The main criteria for comparison are accuracy and bias of the evaluations. Two basic methods have been applied in the literature for evaluation of genomic breeding values. In the first method, applied by Meuwissen et al. [37], simulated data sets are derived, and genomic breeding values are computed on the simulated data. This method has the advantages that the estimated breeding values can be compared to the "true" (simulated) breeding values, and that any number of simulated data sets can be generated. The main disadvantage is that it is not known how accurately the simulation algorithm actually corresponds to actual data.

The second method, first applied by [15], is based on analysis of actual data. On real data the true breeding values are not known. Genomic estimated genetic values for young animals with genotypes, but without trait records or progeny records, are compared to standard estimated breeding values on the same animals based on progeny records produced later. The data is divided into "training" and "validation" data sets. Generally, this is accomplished by dividing the population into old and young animals, respectively. If the validation set is comprised of young animals, only animals with their own trait records or records of progeny can be used for validation. In this case, the squared correlation between the two estimates should equal the reliability of the genomic evaluations divided by the reliability of the daughter-based evaluations [15]. Bias of the evaluation is measured by the regression of true genetic values on estimated genetic values. If evaluations are unbiased then this regression should not be significantly different from unity. That is the covariance of the estimated and true genetic values should equal the variance of the estimated genetic values.

VanRaden et al. [15] used genotypes for 38,416 markers and the August 2003 genetic evaluations for 3576 Holstein bulls born before 1999 to predict the January 2008 daughter deviations for 1759 bulls born from 1999 through 2002. Five milk yield traits, 5 fitness traits, 16 conformation traits, and net merit were analyzed. Combined predictions were more accurate than official PA for all 27 traits, with coefficients of determination 0.05 to 0.38 greater for the genomic evaluations. Reduction of the number of markers by half had virtually no effect on the coefficients of determination for the genomic evaluations. Over the range of 1151–3576 predictor bulls, gains in reliability for net merit were nearly linear with increasing numbers of predictor bulls, and gains for most other

individual traits followed that same pattern. More recent results show that this trend appears to hold up to 10,000 predictor bulls [35]. Numerous studies have shown that accuracies of genomic evaluations are no higher than PA, if less than 1000 bulls are included in the training population (e. g., [15, 21]). Colombani et al. [38] found that regression slopes of observed DYD on predicted DYD for Holsteins were less than unity for all methods for all traits, but highest for standard mixed model evaluations. On a practical level this means that the highest genomic evaluations are somewhat inflated. Differences in accuracy between single and multi-stage methodology were minimal, but single-stage methodologies might have an advantage in bias, and for genetic evaluation of young animals that were not genotyped [29].

8 How Genomics Has Changed the World Dairy Cattle Industry

The main objective for the establishment of the International Bull Evaluation Service (Interbull) was to compute across country genetic evaluations. The Multiple Across Country Evaluation (MACE, [39]) algorithm was developed for this purpose, based on bulls with genetic evaluations in more than a single country (http://www.interbull.org/ib/interbullactivities). The major differences in methodology, and the fact that genomic evaluations of young sires still have low reliability relative to progeny-tested bulls, make the question of across country comparisons of genomic evaluations even more difficult. New methodologies were developed by VanRaden and Sullivan [40] for this purpose, and were incorporated into the Genomic MACE, "GMACE," algorithm that is used by Interbull to compute multi-country genomic evaluations.

Reduction in genotyping costs led to an exponential increase in the number of animals genotyped per year. By March 2021, over 3.8 million Holsteins have been genotyped in North America, of which over 90% are females (https://queries.uscdcb.com/Genotype/cur_density.html). Costs for lower density chips have reached the level at which routine genotyping of all female calves can be economically justified for management decisions on the farm level [41, 42].

Until the advent of genomic evaluation, gains in efficiency of breeding programs due to increase in scale were minimal above population sizes of ~100,000 milk-recorded cows. Thus, many countries were able to run national breeding programs, and the largest countries were also able to conduct breeding programs for more than a single breed. With genomic evaluation, there is a consensus that the main factor limiting the accuracy of genomic evaluations is the number of bulls with genotypes and daughter records (e.g., [15]). Thus, even the largest countries are pooling

resources to conduct multi-national genomic selection programs. Two major consortiums have been established; the Council on Dairy Cattle Breeding (CDCB), which includes organizations from the US, Canada, UK, Italy, Switzerland, and Japan (https://www.uscdcb.com/); and Eurogenomics, which includes members from Spain, the Netherlands, Belgium, France, Germany, Poland, Denmark, Sweden, and Finland (http://www.eurogenomics.com/). Eurogenomics currently included 35,000 bulls and 1.6 million animals in the reference population. Membership in the Eurogenomics consortium is dependent on submission of a quota of bulls with genotypes and genetic evaluations based on daughter records. Recently a consortium of smaller countries that do not meet the requirements of admission into Eurogenomics, "InterGenomics-Holstein," has been organized by Interbull to pull genotypes and compute genomic evaluations for the smaller countries. Breeding organizations from four countries are preliminary members: Ireland, Israel, Slovenia, and South Korea [43].

In the progeny test scheme, described in Fig. 1, a second crop of daughters was produced from elite bulls approximately 5 years after the first crop. This is no longer the case for genomic breeding programs. An example for the Israeli Holstein genomic breeding program as optimized by Reiner-Benaim et al. [44] is diagramed in Fig. 4. The number of bull calves genotyped is 400, as compared to 50 bull calves progeny tested in the scheme in Fig. 1, but 40 are used as sires each year. Thus, selection intensity along the sire-to-sire path is still 1 in 10, but the mean generation interval along the sire-to-dam path is reduced from ~7 to ~2 years. With genomic selection, most bulls are used for general service for less than 1 year, and only a very small sample of elite bulls are retained for further service. The lack of ties between bulls of different ages will make it more difficult to obtain accurate genetic evaluations across generations [46].

With genomic evaluation, progeny testing of large numbers of bulls is no longer necessary. Weller et al. [47] showed that for six countries, US, Canada, Germany, France, Netherlands, and Israel; the number of bulls with daughter records has been reduced from ~4500 in 1995 to ~2500 in 2010. Although the number of US bulls with daughters decreased from ~1500 to ~1100; for Germany, France, and the Netherlands the number of bulls with daughters was reduced by more than half. The decrease was most dramatic in Germany, which progeny tested >1000 bulls born in 1996, but only 308 in 2011. Although the progeny test of young bulls with 50–100 daughters has not been completely eliminated, numbers of bulls have decreased, and the mean number of daughters per young bull has increased.

Daughters are produced only from bull calves with high genomic evaluations. These bulls are a selected sample with respect to the "Mendelian sampling" component of the genetic variance of

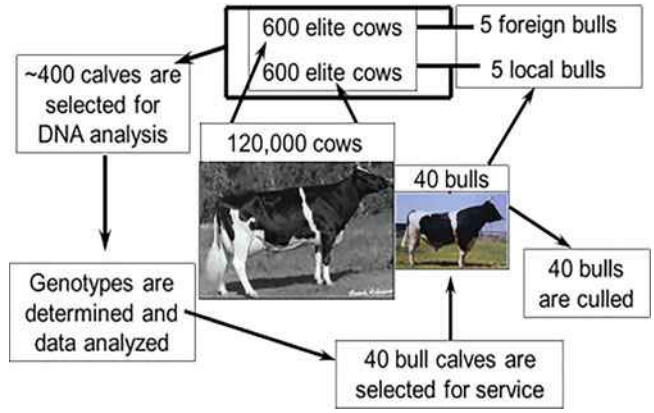

Fig. 4 The Israeli dairy cattle breeding program with genomics as optimized based on [44]. Currently genomic evaluations of the bull calves are computed by CRV as described in [45]

their parents. The genetic value of an individual can be decomposed into the contribution from its sire, its dam and a Mendelian sampling component, which should account for half of the genetic variance in the population. However, if bull calves are selected based on genotype, then only individuals with high Mendelian sampling effects will be selected, which will bias evaluations based on the standard relationship matrix [36]. Selection of bulls out of an already selected population can also result in bias in the calculated genetic gain [42].

García-Ruiz et al. [48] analyzed data from the US Holstein population after nearly 2 generations of genomic selection. They found major reductions in generation intervals, especially along the sire of bull and sire of cow paths. Generation intervals along these paths were reduced from 7 to 2.4 and 5 years, respectively. The change in the dam of bull interval was smaller, from 5.2 to 2.6 years, even though this was originally assumed to be the main pathway to increase genetic gain via genomic selection [1]. More recent data from has shown a further reduction to generation intervals of ~2.3 year for the sire and dam of bulls paths, and 3.8 years for sire of cows for offspring born in 2918 [49]. The number of new artificial insemination bulls produced per year had decreased. Annual rates of genetic gain increased from ~50 to 100% for yield traits and from threefold to fourfold for lowly heritable traits, including female fertility, herd-life and somatic cell concentration [48]. It should be noted though, that these calculations are based on the genomic evaluations, and generally have not been validated by bulls with evaluations based on daughter records. The major engine of gain was the use of much younger bulls as bull sires.

9 The Future

Genetic theory suggests that at some point genetic progress should plateau either through exhaustion of genetic variability or through development of antagonistic genetic relationships between the selection objective and components of fitness. In long-term selection experiments, selection response usually ends after 20–30 generations [50], although in some cases significant response has continued for over 100 generations [51]. Despite these considerations, heritability of lactation milk yield in dairy cattle has actually risen from ~25% in the 1950s to ~35% currently, likely partly due to improved management [52]. Makanjuola et al. [49] compared rates of inbreeding in the US Holstein and Jersey populations during the last decade to the 1990s, which was prior to genomic selection. Per generation, genomic inbreeding rates were slightly higher for Holsteins, but lower for Jerseys. Apparently, little variation has been lost, and current rates of genetic gain are indeed sustainable in the future. Genetic variance for economic traits is maintained by increase in frequency of rare alleles, new mutations, and changes in selection goals and management. As shown by Weller et al. [47], selection can also increase genetic variance if the frequency of the favorable allele is low.

Probably the most important challenge in the future will be incorporation of new traits into the selection index, including udder health, hoof health, other health traits, feed efficiency and methane emissions [53–56]. Although there is consensus as to the economic importance of most of these traits, they have not generally been included in breeding programs because of difficulty of measurement and low heritability for nearly all health traits.

Throughout the chapter, we analyzed genomic evaluation as applied within a single breed. Although crossbreeding in dairy cattle is less common than other agricultural species, such as poultry and swine, there are commercial dairy crossbreeding systems, especially in the tropics, and genomic evaluation has also been applied to crossbred populations. Almost 50% of the New Zealand dairy cattle population are crossbred cows [57]. Khansefid et al. [57] found that crossbred predictions derived from separate analyses of New Zealand Holstein and Jersey reference populations were similar in accuracy to crossbred predictions derived from the two separate purebred reference sets and combined proportional to breed composition. However, the latter approach, which included crossbred animals, was less biased by 0.13.

The main opportunity that has yet to be exploited with respect to genomic evaluation is the huge number of genotyped cows. As noted previously, over 3.5 million cows have been genotyped in North America. Various studies have proposed deriving genomic

evaluations for small populations based on genotypes of cows (e. g., [58, 59]). Pryce et al. [60] estimated that reliabilities of genomic evaluations could be increased by up to 7%, with the incorporation of 10,000 cow genotypes in addition to 3000 bull genotypes. For a trait with heritability of 0.30 and cows with single records, <10 genotyped cows yield equivalent information to a single progeny-tested bull with a reliability of 0.8 [61].

With genotyping costs of <50$ per animal, and accurate imputation techniques, Weller et al. [44] predicted that future genomic evaluation algorithms will be based on a single-step methodologies incorporating hundreds of thousands of cows with genotypes and trait records (e. g., [62, 63]). As noted previously, Misztal and Legarra [33] found problems with convergence of solutions in single-step methodologies. However, in all actual applications of single-step methodology to date only a very small fraction of the animals included in the analysis actually have genotypes. It is likely that convergence problems will decrease if a larger fraction of the population is genotyped. Since cows generally produce from three to five lactation records, the current problem of shortage of genetic ties among animals of different ages will also be solved.

10 Conclusions

The infinitesimal model of genetic variation in quantitative traits seems to correspond very closely to reality. Genomic selection in dairy cattle has led to enhanced rates of genetic improvement, shortening of generation intervals, reduction in the number of bulls with daughter records, and less genetic ties across years. In the future, more use will be made of cow genotypes, with the adoption of single-step methodologies for genomic evaluation. More emphasis will be placed on health, reproduction, efficiency of production, and environmentally friendly production. Genetic variance for economic traits is maintained by increase in frequency of rare alleles, new mutations, and changes in selection goals and management. Thus, it is unlikely that a selection plateau will be reached in the foreseeable future.

References

1. Weller JI (2009) Quantitative trait loci analysis in animals, 2nd edn. CABI Publishing, London, p 272
2. Hayes B, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol 33: 209–229
3. Bennewitz J, Reinsch N, Reinhardt F, Liu Z, Kalm E (2004) Top down preselection using marker assisted estimates of breeding values in dairy cattle. J Anim Breed Genet 121:307–318
4. Boichard D, Fritz S, Rossignol MN, Guillaume F, Colleau JJ, Druet T (2006) Implementation of marker-assisted selection: practical lessons from dairy cattle. In: Proc 8th world cong genet appl livest prod. Belo Horizonte, MG, Brazil, 22 Nov 2006

5. Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD et al (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5:247–252
6. Silva MVB, dos Santos DJA, Boison SA, Utsunomiya ATH, Carmo AS et al (2014) The development of genomics applied to dairy breeding. Livest Sci 166(SI):66–75
7. Henderson CR (1976) A simple method for the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83
8. Thompson R (1979) Sire evaluations. Biometrics 35:339–353
9. Westell RA, Quaas RL, Van Vleck LD (1988) Genetic groups in an animal model. J Dairy Sci 71:1310–1318
10. VanRaden PM, Wiggans GR (1991) Derivation, calculation and use of national animal model information. J Dairy Sci 74:2737–2746
11. Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388
12. Weller JI, Kashi Y, Soller M (1990) Power of "daughter" and "granddaughter" designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. J Dairy Sci 73: 2525–2537
13. Ron M, Band M, Yanai A, Weller JI (1994) Mapping quantitative trait loci with DNA microsatellites in a commercial dairy cattle population. Anim Genet 25:259–264
14. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R et al (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907–920
15. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24
16. Fernando RL, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467–477
17. Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Proc Roy Soc Edinburgh 52: 399–433
18. Maher B (2008) Personal genomes: the case of the missing heritability. Nature 456:18–21
19. Visscher PM (2008) Sizing up human height variation. Nat Genet 40:489–490
20. VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91: 4414–4423
21. Weller JI, Glick G, Shirak A, Ezra E, Seroussi E et al (2014) Predictive ability of selected subsets of single nucleotide polymorphisms (SNPs) in a moderately sized dairy cattle population. Animal 8:208–216
22. Ron M, Weller JI (2007) From QTL to QTN identification in livestock—"Winning by points rather than knock-out": a review. Anim Genet 38:429–439
23. Weller JI, Ron M (2011) Invited review: quantitative trait nucleotide determination in the era of genomic selection. J Dairy Sci 94: 1082–1090
24. Grisart B, Coppieters W, Farnir F, Karim L, Ford C et al (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res 12:222–231
25. Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A et al (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res 15:936–944
26. Zhi-Liang H, Park CA, Reecy JM (2019) Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Res 47:D701–D710
27. Weller JI, Bickhart DM, Wiggans GR, Tooker ME, O'Connell JR et al (2018) Determination of quantitative trait nucleotides by concordance analysis between quantitative trait loci and marker genotypes of US Holsteins. J Dairy Sci 101:9089–9107. https://doi.org/10.3168/jds2018-14816
28. Ron M, Cohen-Zinder M, Peter C, Weller JI, Erhardt G (2006) ABCG2 polymorphism in Bos indicus and Bos taurus cattle breeds. J Dairy Sci 89:4921–4923
29. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci 93:743–752
30. Weller JI, Ezra E (2004) Genetic analysis of the Israeli Holstein dairy cattle population for production and non-production traits with a multitrait animal model. J Dairy Sci 87:1519–1527
31. Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and

genomic information. J Dairy Sci 92: 4656–4663
32. Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92:4648–4655
33. Misztal I, Legarra A (2017) Invited review: efficient computation strategies in genomic selection. Animal 11:731–736
34. Wiggans GR, VanRaden PM, Cooper TA (2012) Technical note: adjustment of all cow evaluations for yield traits to be comparable with bull evaluations. J Dairy Sci 95: 3444–3447
35. Weller JI (2016) Genomic selection in animals. Wiley, Hoboken, NJ, p 175
36. Patry C, Ducrocq V (2010) An approach to account for selection bias in national evaluations due to genomic selection. In: Proc 9th world cong genet appl livest prod 0559 Leipzig, Germany
37. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
38. Colombani C, Legarra A, Fritz S, Guillaume F, Croiseau P et al (2013) Application of Bayesian least absolute shrinkage and selection operator (LASSO) and Bayes Cπ methods for genomic selection in French Holstein and Montbeliarde breeds. J Dairy Sci 96:575–591
39. Schaeffer LR (1994) Multiple-country comparison of dairy sires. J Dairy Sci 77: 2671–2678
40. VanRaden PM, Sullivan PG (2010) International genomic evaluation methods for dairy cattle. Genet Sel Evol 42:7
41. Thomasen JR, Willam A, Guldbrandtsen B, Lund MS, Sørensen AC (2014) Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. J Dairy Sci 97:458–470
42. Calus MP, Bijma P, Veerkamp RF (2015) Evaluation of genomic selection for replacement strategies using selection index theory. J Dairy Sci 98:4945–4955
43. Klopčič M (2018) Introduction of InterGenomics Holstein session. Interbull web publication. https://interbull.org/static/web/IgHOL_Intro.pdf
44. Reiner-Benaim A, Ezra E, Weller JI (2017) Optimization of a genomic breeding program for a moderately sized dairy cattle population. J Dairy Sci 100:2892–2904
45. Weller JI, Stoop WM, Eding H, Schrooten C, Ezra E (2015) Genomic evaluation of a relatively small dairy cattle population by combination with a larger population. J Dairy Sci 98:4945–4955
46. Liu Z, Alkhoder H, Reinhardt F, Reents R (2016) Accuracy and bias of genomic prediction for second-generation candidates. In: Proc interbull meet Puerto Varas, Chile, 24–28 Oct 2016
47. Weller JI, Ezra E, Ron M (2017) Invited review: a perspective on the future of genomic selection in dairy cattle. J Dairy Sci 100: 8633–8644. https://doi.org/10.3168/jds2017-12879
48. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci 113:E3995–E4004
49. Makanjuola BO, Miglior F, Abdalla EA, Maltecca C, Schenkel FS, Baes CF (2020) Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J Dairy Sci 103:5183–5199
50. Kennedy BW (1984) Selection limits: have they been reached with the dairy Cow? Can J Anim Sci 64:207–215
51. Hill WG (2008) Estimation, effectiveness and opportunities of long term genetic improvement in animals and maize. Lohmann Inf 43: 3–20
52. Hill WG (2016) Is continued genetic improvement of livestock sustainable? Genetics 202: 877–881
53. Egger-Danner C, Cole JB, Pryce JE, Gengler N, Heringstad B et al (2015) Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal 9:191–207
54. Chesnais JP, Cooper TA, Wiggans GR, Sargolzaei M, Pryce JE, Miglior F (2016) Using genomics to enhance selection of novel traits in North American dairy cattle. J Dairy Sci 99:2413–2427
55. Schöpke K, Swalve HH (2016) Review: opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal 10: 1050–1060
56. de Haas Y, Pszczola M, Soyeurt H, Wall E, Lassen J (2017) Invited review: phenotypes to genetically reduce greenhouse gas emissions in dairying. J Dairy Sci 100:855–870
57. Khansefid M, Goddard ME, Haile-Mariam M, Konstantinov KV, Schrooten C, de Jong G, Jewell EG, O'Connor E, Pryce JE, Daetwyler HD, MacLeod IM (2020) Improving genomic prediction of crossbred and purebred dairy

cattle. Front Genet 11:598580. https://doi.org/10.3389/fgene.2020.598580

58. Jenko J, Wiggans GR, Cooper TA, Eaglen SAE, de L Luff WG et al (2017) Cow genotyping strategies for genomic selection in a small dairy cattle population. J Dairy Sci 100: 439–452
59. Uemoto Y, Osawa T, Saburi J (2017) Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. Animal 11:382–393
60. Pryce JE, Hayes BJ, Goddard ME (2012) Genotyping dairy females can improve the reliability of genomic selection for young bulls and heifers and provide farmers with new management tools. Proc 38th ICAR session 28
61. Boichard D, Ducrocq V, Fritz S (2015) Sustainable dairy cattle selection in the genomic era. Invited review. J Anim Breed Genet 132: 135–143
62. Koivula M, Strandén I, Aamand GP, Mäntysaari EA (2016) Effect of cow reference group on validation reliability of genomic evaluation. Animal 10:1061–1066
63. Mäntysaari EA, Strandén I (2016) Genomic data and breeding value estimation in dairy cattle: theory, practice, problems. J Anim Breed Genet 133:165–166