

**Accomplishments and new challenges in dairy genetic evaluations**

Christian Maltecca,1 Kristen L. Parker,1 Martino Cassandro2

1Department of Animal Science, North Carolina State University, Raleigh, NC, USA
2Dipartimento di Scienze Animali, Università di Padova, Italy

**Abstract**

This review presents the evolution of dairy genetic methods to estimate breeding values. For centuries, human action has shaped animal populations by choosing progenitors of the next generation. Since the twentieth century, applied concepts were integrated into a new discipline, quantitative genetics. The past quarter-century in genetic evaluation of dairy cattle has been marked by evolution in methodology and computer capacity, expansion in the array of evaluated traits, and globalization. Selection index was replaced by mixed model procedures and animal models replaced sire and sire-maternal grandsire models. Recently, application of Bayesian theory to breeding values prediction and variance components estimation has become standard. Individual test-day observations have been used more effectively in the estimation of lactation yield as direct input to evaluation models. Computer speed and storage are less limiting in choosing procedures. National evaluations combined internationally provide evaluations for bulls from all participating countries on each of the national scales, facilitating choices from among many more bulls. Selection within countries has increased inbreeding and the use of similar genetics across countries reduces the previously available genetic diversity. Finally, considerable progress in genomics has created a new tool, genomic selection. The collection and analysis of several types of phenotypic data to evaluate genetic merit will continue to be the most important tool for genetic progress in the foreseeable future. Information will increasingly be obtained from smaller reference populations and the extrapolation from these data will require careful validation.

**Introduction**

Humans have recognized the utility of artificial selection long before formalizing it. For centuries, human action has shaped animal populations by choosing progenitors of the next generation, whether solely based on their phenotypes, as probably during the process of domestication, or with the more or less explicit understanding that resemblance among related individuals can be used as a selection tool. At the dawn of the twentieth century, most of these concepts were formalized into a new discipline, called quantitative genetics. From then on, improvement of quantitative traits in livestock relied on the accurate recording of phenotypic records of individuals and their relatives, in addition to the knowledge of the genetic variation for the traits. This approach, albeit successful, proved relatively slow for traits measured late in life, traits that are difficult and expensive to measure, or sex-limited traits. Traditional selection schemes have relied mainly on the male path of selection. Progeny testing of sires is a long and costly process taking about 5 years and approximately \$50,000 per bull (Schaeffer, 2006). Advances in molecular genetics in the second half of the last century promised to expand, or even reshape, the landscape of livestock selection after recombinant DNA technologies led to the identification of DNA polymorphisms, and eventually to the development of genetic linkage maps for several livestock species. In the 1990s, interest piqued in the identification of quantitative trait loci (QTL) affecting economically important traits in livestock. For the vast majority of cases however, the direct utilization of genotypic information from these findings in selection programs through marker-assisted selection (MAS) proved far more challenging than anticipated (Dekkers and Hospital, 2002; Dekkers, 2004). Dense panels of single nucleotide polymorphisms (SNP) are now available for most domestic animal species. The integration of this large amount of molecular data with phenotypic and ancestry information opened new and fascinating scenarios, while at the same time posing new challenges in livestock selection.

**Quantitative genetics as a science**

After the rediscovery of Mendel’s work in the first years of the 20th century, a bitter controversy began with Mendelian geneticists, such as William Bateson, versus Biometricians, such as Karl Pearson (Lynch and Walsh, 1998). The inheritance of discrete characters seemed somehow irreconcilable with the concept of continuously distributed traits. This dichotomy had relevant reflections, especially on evolutionary biology. Was selection acting through mutations with large effects or was it shaped by continually distributed traits? The question was (partially) resolved by the work of R.A. Fisher (1918), who posed the basis for a unified science. Fisher postulated that a large number of Mendelian factors jointly control phenotypic characters, each contributing a small amount to the overall phenotypic variance. Under these conditions, the inheritance of continuous variation can be explained within the Mendelian framework. An important consequence of Fisher’s model is the possibility of deriving genetic correlations between individuals, thus allowing the practical implementation of artificial selection. The formalization of several concepts related to this last point is to be credited to S. Wright, arguably the founding father of both quantitative genetics and modern animal breeding. In a series of papers known as “System of matings” published by Wright (1921a; 1921b; 1921c; 1921d; 1921e), the author derived and defined the application of essential concepts such as relationship, inbreeding, and heterozygosity, along with quantifying the effects of different types of matings. In addition, Wright introduced the idea of path coefficients and of path analysis, which will be later employed in the derivation of the selection index.

The introduction of these new theoretical developments led to the formalization of animal breeding as a science in the mid of the 20th century, mainly due to the work of the Iowa group, led by J.L. Lush and L.N. Hazel. The Iowa group represented a bridge from the...
Theoretical advancement to the practical implementation of selection programs. Lush and Hazel combined both path analysis and correlation and regression methods, blending the use of family and individual performances in genetic improvement, developing selection index theory. Furthermore, the Iowa group understood the importance and fostered the development of precise measures and recording of phenotypes in selected populations.

The selection index

A selection index in its simplest form is a prediction of animals’ breeding merits for total economic merit. Total economic merit is a combination of an animal’s breeding value for all economically important traits, with each trait weighted by its net economic value (Shook, 2006). Index selection was developed independently by Smith (1936) using Fisher’s discriminant function, and by Hazel (1943) using multiple regression and path coefficient analysis. A selection index allows maximizing selection response in one or more traits by using several sources of information on relatives, including the individual’s own record. Although selection index theory relies on Fisher’s infinitesimal model, it can be applied to characters regulated by few genes under the assumption of additivity. Since most selection schemes involve the improvement of multiple traits simultaneously, it is convenient to define an aggregated breeding value as:

\[
H = a_1B_{V1} + a_2B_{V2} + \ldots + a_nB_{Vn}
\]

where each trait i differs in economic value as measured by ai, and BVi represents the breeding value for each specific trait entering the aggregated BV.

The index (I) used to predict the estimated breeding value of an individual (H) can be expressed as:

\[
I = b_1P_1 + b_2P_2 + \ldots + b_nP_n
\]

where: b, are index weights and Pi, are phenotypic deviations from population mean. Using matrix notation:

\[
a = n \times 1 \text{ column vector of known economic values,}
\]

\[
b = m \times 1 \text{ column vector of partial regression coefficients to be solved for,}
\]

\[
P = m \times m \text{ phenotypic variance-covariance matrix,}
\]

\[
A = n \times n \text{ additive genetic variance covariance matrix,}
\]

\[
C = m \times n \text{ matrix of covariances between phenotypic values in I and additive genetic values in H.}
\]

Genetic gain in H given I can be expressed as:

\[
\Delta G_{HI} = b_H^T \sqrt{V_I} = iR_{HI} \sqrt{V_H} = iR_{HI} \sqrt{V_H} \quad \text{(3)}
\]

where:

- i is the selection intensity,
- BH is the regression of H on I,
- RH is the correlation between H and I, or accuracy,
- Vf is the variance of the index = b'Tb,
- and CHI is the covariance between H and I = b'C.

The index coefficients are derived by differentiating \(\Delta G_{HI}\) (assuming i constant) with respect to b, setting the simultaneous equations equal to zero and solving for b:

\[
\frac{\partial \Delta G_{HI}}{\partial b} = - \frac{Ca}{(b'Pb)^2} - \frac{2 \text{a'Ca}(Pb)}{(b'Pb)^2} = 0
\]

\[
\frac{Ca}{(b'Pb)^2} (b'Pb) = 0
\]

\[
Ca = Pb = P^{-1}Ca \quad \text{(4)}
\]

where \(P^{-1}\) is the inverse of P.

Rönningen (1985) listed the properties of a selection index as follows:

i. If H is the true breeding value, then RH is maximized;

ii. \(E(H-I)^2 = (1-R^2) VH\), the variance of prediction errors, is minimal among all linear functions of the general form of the selection index;

iii. The probability of selecting one of the largest sample values of total merit is maximal by selecting the largest value of the index criteria;

iv. The probability of selecting the higher merit of any two individuals is maximal;

v. The genetic progress in any one round of selection by the index is maximal.

The selection index assumes the exact knowledge of the genetic parameters, which is seldom the case. In practice, the index appears to be robust so that moderate errors in estimation of genetic parameters do not severely compromise use of the index approach (Sales and Hill, 1976). Furthermore, it should be noted that the index assumes knowledge of all the fixed effects, which again is seldom guaranteed. Linearity in the economic values is also assumed, although a relaxation of this assumption has been discussed in Wilton and Van Vleck (1968). Lastly, a multivariate normal distribution (MVN) of the traits is assumed, which cannot always be guaranteed.

Various modifications of the selection index have been proposed; a complete review is beyond the scope of this paper but can be found in Lin (1978) and Baker (1986). Cunningham (1969) showed that a selection index could be reduced relative to a complete index by eliminating one or more sources of information, leading to a reduced index. Williams (1962a, 1962b) developed the base index, making the application of selection index possible when data is not available to estimate the genetic and phenotypic parameters for the weighting factors in a conventional index. Elston (1963) developed the weight free index, which does not require the estimation of parameters and economic weights. The phenotypic covariate index was suggested by Rendell (1954) and discussed further by Searle (1965) for situations in which breeders may choose to use standard covariance analyses to correct trait 1 for trait 2 phenotypically. The desired gains index, was introduced by Pesek and Baker (1969) and expanded by several authors (Tallis, 1962; James, 1968; Harville, 1975; Yamada et al., 1975). This index allows breeders to capitalize on a notion of the amount of response a set of traits should have relative to each other, without having any basis for assigning economic values to the traits. A general case occurs when the goal is to maximize H for n traits and provide relative desired gains \(d_i\) for \(n\) other traits. The problem was solved by Harville (1975), who developed the selection index with general constraints. Finally, Lande and Thompson (Lande and Thompson, 1990) derived selection indices that maximize the response to selection using phenotypic traits and genetic markers in a combined index.

Linear mixed models and MME equations

In 1949, C.R. Henderson, a student of L.N. Hazel, derived the mixed model procedure, which is arguably the most important methodological development in the field of biometrics applied to animal breeding. Henderson generalized and proved properties of the mixed model procedure in the course of the following twenty years (Henderson, 1975; 1984). The mixed model procedure combines best linear unbiased prediction (BLUP) of random effects, such as additive genetic value and permanent environmental effects, with best linear unbiased estimates (BLUE) of fixed factors. Best in
this case being defined as minimizing the variance of prediction error for procedures that are unbiased among linear functions of the data.

Mixed model methodology provides a flexible, versatile, and often computable statistical tool for enhancing productivity of livestock. Indeed, production testing often occurs in a range of environments, e.g., herds with heterogeneous genetic and residual variances, as well as covariances. In this circumstance, across-herd selection can be viewed as a problem of choosing candidates from several distributions.

The mixed model can be described in matrix notation terms as follows:

\[ y = Xb + Zu + e \]  \[ 5 \]

where \( y \) is an \( N \times 1 \) vector of phenotypic observations, \( X \) is the incidence matrix of fixed effects, \( b \) is the \( p \times 1 \) vector of fixed effects, \( Z \) is the incidence matrix of observations on the animal, \( u \) is the \( q \times 1 \) vector of random effects, and \( e \) is the \( N \times 1 \) random vector of residuals. Expectations of the model are as follows:

\[ E(u) = 0 \]
\[ E(e) = 0 \]
\[ E(y) = E(Xb + Zu + e) = Xb \]

Variance covariance structure is normally represented as

\[ \nu(u,e) = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix} \]

with \( G \) and \( R \) being known positive definite matrices. As a result:

\[ \text{Var}(y) = \text{Var}(Xb + Zu + e) = Z\text{Var}(u)Z' + \text{Var}(e) + Z\text{Cov}(u,e) + \text{Cov}(e,u)Z' = ZGZ' + R \]  \[ 6 \]

where \( G \) is the additive genetic variance and in the univariate case equal to \( A^*\sigma^2_a \); \( A \) represents the matrix of kinship between animals and \( \sigma^2_a \) is the additive genetic variance of character known a priori. \( R \) represents the residual variance matrix and in its simplest form for a univariate model reduces to \( 1^*\sigma^2_e \), where \( 1 \) is an identity matrix and \( \sigma^2_e \) is the residual variance.

Solutions for \( b \) (BLUE) and \( u \) (BLUP) can be obtained using general prediction theory. Since the focus of this paper is in breeding values prediction, only BLUP for the random effects will be derived. The general procedure for obtaining BLUE is not dissimilar and is left to the reader.

Provided a \( Kb \) estimable function we wish to predict:

\[ Kb + Mu \]

The predictor is the function to predict through, a linear function of \( y \), i.e. \( Ly \), for some \( L \).

By restricting the search to the class of linear and unbiased, the best predictor has form:

\[ K\hat{b} + CV^{-1}(y,\hat{b}) \]

where

\[ \hat{b} = (X'V^{-1}X)^{-1}X'V^{-1}y \]

\[ V \]

is as previously defined, \( C = \text{cov}(Kb + Mu) \), and \( \hat{b} \) can replace \( b \) in the predictor’s formula.

Because the predictor is to be unbiased, the mean squared error is equivalent to the variance of prediction error and the function to be minimized \( F \) can be written as:

\[ F = \text{Var}(PE) + (L'X - K')\Phi \]

with \( \Phi \) representing a Lagrange multiplier to force unbiasedness. Differentiating \( F \) with respect to the unknowns \( L \) and \( \Phi \), and equating the partial derivatives to 0:

\[ \frac{\partial F}{\partial L} = 2VL - 2ZGM + XL = 0 \]
\[ \frac{\partial F}{\partial \Phi} = X'L - K = 0 \]

Letting

\[ \theta = 0.5\Phi \]

the first derivative can be written as:

\[ VL = ZGM - X\theta \]

and solving for \( L \):

\[ V^{-1}VL = L \]
\[ = V^{-1}ZGM - V^{-1}X\theta \]

Substituting the above for \( L \) into the second derivative and solving for 0:

\[ X'K = 0 \]
\[ X'(V^{-1}ZGM - V^{-1}X\theta - K) = 0 \]
\[ X'X\theta = XV^{-1}ZGM - K \]
\[ \theta = (XV^{-1}X)^{-1}(XV^{-1}ZGM - K) \]

Substituting this solution for \( \theta \) into the equation for \( L \) gives:

\[ L = M'GZV^{-1} + K'(XV^{-1}X)^{-1}XV^{-1} - M'GZV^{-1}X(XV^{-1}X)^{-1}XV^{-1} \]

By substituting the BLUE for \( b \) into the formula, the predictor becomes:

\[ Ly = K\hat{b} + MgZV^{-1}(y - \hat{b}) \]

The BLUP predictor for random effects

Since \( V \) is usually too large to be directly inverted the BLUP predictors (along with the BLUE estimators) are obtained starting from Henderson’s mixed model equations (MME) (Kennedy, 1989). These equations are of order equal to the number of elements in \( b \) and \( u \), which are usually smaller than the number of elements in \( y \), making finding the solution easier. Also, these equations require the inverse of \( R \), which has a simpler structure than \( V \) since it is diagonal.

\[ [X' R^{-1} X | X' R^{-1} Z | Z' R^{-1} X | Z' R^{-1} Z + G^{-1} | b | u] = [X' R^{-1} y | Z' R^{-1} y] \]

By comparing selection index theory with the mixed model equation, it is possible to verify that the MME combines least-squares estimation with selection index in order to derive unbiased estimates of genetic values of individuals sampled in different environments, such as herds or blocks. This development allowed the overcoming of one of the main limitations of the index selection: the assumption of a priori knowledge of the true average of the fixed effects.

The properties in common between BLUP and selection index are described by Van Vleck (1993): i) both are unbiased; ii) variances of prediction errors are minimized; iii) the correlation between the prediction and true value is maximized; iv) the predictions maximize the probability of correct ranking; v) the predictions of MME are the same of selection index when fixed effects null a pre-adjusted.

The BLUP method has now become the standard animal genetic evaluation method, though its application could not occur for over 20 years after its implementation due to unavailability of computers that could handle large arrays, especially an inverse \( G \) matrix of large size. Only in the early 1970s was it possible to solve the first models with MME for large datasets, and since then the BLUP model has been used in genetic evaluations for the vast
majority of livestock species. Furthermore, MME have found successful applications in crop science and in other statistical fields.

**Expanding the mixed model**

Several improvements and generalizations have marked the history of the MME, especially focusing on the analysis of a multivariate model, whether through the use of models with multiple traits recorded per animal (multiple traits model) or with repeated observations of the same trait per animal (repeatability models) (Van Vleck, 1993).

MME models have grown in size over time, along with computer capability. In dairy cattle, we transitioned from models that considered only the additive effect of father (sire models) (Van Vleck, 1993) on the observations recorded on daughters, to more comprehensive models such as the maternal grandsire model (Van Vleck, 1993), to the animal model currently used for the vast majority of traits which has been defined as the ideal model in the reality of genetic evaluation in dairy cattle breeding (Quaas and Pollak, 1980). Many applications of the MME procedure assume that variance components are constant across subclasses of effects considered. However, genetic, environmental, and phenotypic variances seen in practice are often heterogeneous (Brotherstone and Hill, 1986). The bias caused by ignoring the heterogeneity of variance is severe, especially in the selection of dams of future sons, because progeny of a dam tend to make records in the same herd (Cumberland et al., 1987). Several procedures for accounting for heterogeneous variances have been presented (Gianola et al., 1992; Meuwissen and Van der Werf, 1993). Among them, the procedure using standardization of records before solving the mixed model equations (i.e., 2-step procedure) may work well when the size of the subclasses is relatively small and have applicability to BLUP with the animal model (Hill, 1984).

The adoption of Test-day (Wheatley and Henderson, 1975) models, allowing all genetic and environmental effects to be estimated directly on a test-day basis (Pukk and Schaeffer, 1993), is perhaps the last major development in the time of MME. The TD model allows more accurate genetic prediction by better accounting for specific test-day environment variability rather than relying on lactation yields as input. A test-day model (TDM) improves the accuracy of genetic evaluation, tends to make records in the same herd, and have applicability to BLUP with the animal model (Van Vleck, 1993).

Test-day (TD) models allow more accurate genetic prediction by better accounting for specific test-day environment variability rather than relying on lactation yields as input. Two distant TD yields can contribute more information than those close and highly correlated. TD data allows the use of information from lactations with long intervals between milk recordings because estimation of yields for unrecorded intervals would not be required. On the other hand, a test-day model cannot overcome the loss in accuracy from fewer TD and allows yields from any combination of TD to be included appropriately (Wiggans and Goddard, 1996). In TDMs, records from individual test days are used to determine lactation production instead of aggregating records (Swalve, 2000).

Over the years, fitness and fertility traits have assumed a larger emphasis in selection programs. The application of Henderson’s mixed model equations has, as a consequence, expanded beyond the realm of linear mixed models to include generalized linear mixed models for binary, ordinal, or count traits. These models account for the distributional nature of discrete data in the modeling of expected responses as a function of risk factors (Tempelman, 1998). Development and popularization of generalized linear mixed models for discrete data in animal breeding is mostly to be credited to D. Gianola and J.L. Foulley (Gianola and Foulley, 1983; Foulley et al., 1983; Gianola and Foulley, 1984), although the same ideas were developed independently by Harville and Mee (Mee and Harville, 1982; Harville and Mee, 1984). Briefly considering the model in [5], the standard distributional assumption for the likelihood linear mixed model case is:

$$N(Xb+Zu,R)$$

where R can assume different forms but in the simplest case is $R=I\cdot Fe$. This Gaussian density employed in BLUP models does not fit the analysis of categorical data, and other sampling distributions of the exponential family might be considered. It is nonetheless possible to re-parametrize the model so that the expectation of y can be expressed as (Tempelman, 1998)

$$E(y|\eta)=l(\eta)$$

where $\eta = Xb + Zu$ and l(\eta) represents a link function for which most categorical data are represented by either probit or logit functions. The general linear mixed model can be seen as a special case of a generalized linear model where the link function reduces to an identity function. Threshold mixed models for the re-parametrized models can then be defined and solved for genetic analysis (Foulley et al., 1983). Over the years, threshold models have been expanded to include count (Djennal et al., 1987) and censored data (Gonzalez-Recio et al., 2006). In addition, the application of Weibull mixed models is today largely employed in survival analysis (Ducrocq et al., 1988a; 1988b).

**Bayesian methods**

The growth, both in terms of size and complexity of the models employed in animal breeding presented for a long time computational and theoretical challenges in terms of the solution of large sets of equations and the estimation of variance components. Methods employed throughout the seventies and eighties (mostly based on Maximum Likelihood and Restricted Maximum likelihood) were in some cases ad hoc, and in most cases had some undesirable theoretical consequences. In the mid eighties, thanks to the pioneering work of D. Gianola, a new Bayesian theoretical framework was introduced, which provided an extremely powerful conceptual strategy to solve problems in animal breeding theory (Gianola and Fernando, 1986). Although we will not provide formal treatment of the topic, we will illustrate the basic concepts. For an in depth treatment the reader is referred to Sorensen and Gianola (2002). Bayesian inference makes use of important consequences of a simple probability rule. Consider the application of Bayes theorem to a general case:

$$p(\theta|y) = \frac{p(y|\theta) p(\theta)}{p(y)}$$

Where $p(\theta|y)$ is the joint distribution of $\theta$, the unknowns, and y, our data. Our interest is in obtaining knowledge about the unknowns in the model. We can see that the probability of the unknowns given the data is:

$$p(\theta|y) \propto p(y|\theta) p(\theta)$$

[15] represents the posterior probability function, which is proportional to $p(y)$, the prior probability of the unknown reflecting our prior knowledge of the data and to $p(y|\theta)$, the likelihood, the information obtained from our data.
Let’s once again refer to the general settings in [9]. In the usual animal model, we can show that $p(y | \theta)$ has the form (Gianola and Fernando, 1986):

$$y | \theta, u, \alpha_2^2, \delta_2^2 \sim N(Xb + Zu, R)$$

$$p(y | \theta, u, \alpha_2^2, \delta_2^2) = \frac{1}{(2\pi)^{N/2} | \Omega |^{1/2}} e^{-\frac{1}{2} \frac{1}{\delta_2^2} e^{-\frac{1}{2} (y - Xb - Zu)^T \Omega^{-1} (y - Xb - Zu)}}$$

and $p(\theta)$ is:

$$p(\theta) = p(b)p(u)p(\alpha_2^2)p(\delta_2^2)$$

for which general prior assumptions are:

$$p(b) \propto \text{constant}$$

$$p(u) \propto (\alpha_2^2)^{-\frac{q}{2}} e^{-\frac{1}{2} (uA^{-1}u)/\sigma_u^2}$$

$$p(\alpha_2^2) \propto \text{scale} \ln \chi^2 - \chi^2$$

$$\left(\frac{\alpha_2^2}{2 \sigma_2^2}\right)^{-\frac{q}{2} + 1} e^{-\frac{1}{2} \frac{1}{\sigma_2^2} (uA^{-1}u + \sigma_2^2)\sigma_2^2}$$

where $\nu_u$ and $\sigma_2^2$ are hyperparameters $\sigma_2^2 = (uA^{-1}u)\nu_u$ (de los Campos et al., 2009) and $\nu_u = d.f.$

Similarly,

$$p(\sigma_2^2) \propto \text{scale} \ln \chi^2 - \chi^2$$

$$\left(\frac{\sigma_2^2}{2 \sigma_2^2}\right)^{-\frac{q}{2} + 1} e^{-\frac{1}{2} \frac{1}{\sigma_2^2} (uA^{-1}u + \sigma_2^2)\sigma_2^2}$$

with $\Sigma_2^2 = (\nu\nu^T)/\nu(SSE)$ and $\nu_\nu = d.f.$

Combining [17] and [18-22], it is now possible to obtain the joint posterior distribution:

$$p(\theta | y, \theta) = \frac{p(y | \theta) p(\theta)}{p(y)}$$

$$\propto (\alpha_2^2)^{-\frac{r}{2} + 1} e^{-\frac{1}{2} \alpha_2^2}$$

$$[(y - Xb) - Zu)^T (y - Xb) - Zu + \nu_\nu^2]$$

$$\left(\frac{\alpha_2^2}{2 \sigma_2^2}\right)^{-\frac{q}{2} + 1} e^{-\frac{1}{2} \frac{1}{\sigma_2^2} (uA^{-1}u + \sigma_2^2)\sigma_2^2}$$

$$\left(\frac{\nu_\nu^2}{2 \sigma_2^2}\right)^{-\frac{q}{2} + 1} e^{-\frac{1}{2} \frac{1}{\sigma_2^2} (uA^{-1}u + \sigma_2^2)\sigma_2^2}$$

From this, and making use of conditional probability, it is possible to obtain estimates from the conditional posterior distributions for all the parameters of interest.

Although a full Bayesian approach is appealing, it is also in most cases extremely complex, involving the computation of multidimensional integrals. To obviate this problem, efficient mathematical methods based on simulation have been implemented. Today, Markov Chain Monte Carlo Methods (MCMC) are commonly used in Bayesian analysis. One of them, the Gibbs Sampling (Gelfand et al., 1990), is the method of choice for many genetic analyses because it allows for a large flexibility in the modeling and a straightforward implementation that makes use of the natural hierarchical interpretation of Bayesian models.

**International bull evaluations**

Along with technical advancement in data modelling, probably the most important evolution in genetic evaluations has been the implementation of a sire model for multi-national genetic evaluations for both beef and dairy populations. Indeed, the increasing trade of semen, embryos, and livestock has naturally led breeders towards wanting to make accurate comparisons between animals, primarily bulls, performing both within and across countries. However, these comparisons are difficult due to:

- differences in genetic evaluation methods;
- differences in breeding objectives;
- differences in genetic levels;
- differences in farming environment.

A solution to this problem was first provided using procedures based on linear regression techniques (Weigel, 1997). These procedures include the Wilink method, which uses EBV of bulls with progeny in the importing and exporting countries; the Goddard method, which uses daughter yield deviations of bulls with progeny in the importing and exporting countries; and the full-sib method, which uses daughter yield deviations of pairs of siblings with progeny in either the importing or the exporting country. Limitations of these regression based methods include i) an insufficient number of bulls with progeny in more than one country, ii) instability of conversion equations over time because of changes in the group of bulls used to develop the equations, iii) no possibility for re-ranking of bulls across countries if genotype by environment interaction is present, and iv) reduced accuracy of prediction for elite bulls, which are the bulls of most interest in breeding programs.

An improvement to limitations of linear regression techniques was provided with the so-called MACE (multiple across country evaluation) proposed in 1994 by Larry Schaeffer. The MACE allowed the combination of the genetic evaluation of breeding, developed with the most diverse MME from various countries, into a single international assessment. The MACE is currently in use at the center of Interbull (International Bull Evaluation Service), which is based in Uppsala, Sweden, and consists of 42 member countries that receive quarterly evaluations of all the genetic breeding available on a world ranking. Interbull was developed in 1983 as a joint venture between ICAR, the European Association for Animal Production (EAAP) and the International Dairy Federation (IDF).

However, international comparisons showed some critical aspects due to variability trends over time (Cassandro et al., 1996; Miglior et al., 1997) and selection bias due to deviation from the expected Mendelian sampling value of zero (Fikse, 2004).

**From quantitative trait loci detection to Genomic selection**

Although Fisher’s infinitesimal model served the cause of animal breeding exceptionally well over the last 60 years, it remains an approximation. We do in fact know that there is not an infinite number of unlinked additive loci each contributing a negligible amount to the total variance and that indeed, the amount of genetic material possessed by any individual is finite. The idea that some of the loci responsible for genetic variation in quantitative traits have effects that are large enough to be detected by linkage has long been recognized. The experiments of Sax (1923) with beans demonstrated that the effect of an individual locus on a quantitative trait could be isolated through a series of crosses, resulting in randomization of the genetic background with respect to all genes not linked to the marker under investigation (Weller, 2001). The overall idea of quantitative trait loci (QTL) mapping is to trace chromosome regions associated with phenotypic variation assuming no actual knowledge of the gene(s) influencing the trait. QTL studies provide evidence that a chromosomal region has a certain probability of being heterozygous for one or more QTL that are responsible for a portion of the phenotypic variance. In the 1990s, interest boosted in the identification of QTL affecting economically important traits in livestock and has made possible the identification of certain genes or chromosomal regions associated with phenotypic differences between individuals, families, and breeds. In a few cases, specific genes for which the allelic variant responsible for the phenotypic variation were identified and have been easily incorporated in the estimation of animal breeding values (Visscher, 1996). Unfortunately, these cases represented the exception. The incorporation of QTL in selection schemes has therefore relied on marker-assisted selection (MAS), first implemented within the BLUP general framework by the...
work of Fernando and Grossman (1989). The direct utilization of genotypic information in selection programs through MAS tells a story of mixed results, and although successful in some cases, it required an average of about 25 years for its full implementation (Dekkers, 2004). In the last few years, single nucleotide polymorphisms (SNPs) have somewhat revolutionized the approach to QTL discovery and the overall utilization of molecular markers. A single nucleotide polymorphism is a DNA sequence variation occurring when a single nucleotide in the genome differs between members of a species, or between paired chromosomes in an individual. Although its binary nature makes each single SNP inherently less informative compared to other available markers, SNPs are the densest molecular markers in the mammalian genome. In addition, high-throughput genotyping techniques have drastically decreased the cost of performing large-scale genotyping. Different platforms are now available for parallel high-throughput genotyping of several livestock species. The availability of dense marker panels has made possible the efficient exploitation of population historical linkage disequilibrium as outlined by Meuwissen and Goddard (2000) and has somewhat overcome the low resolution and power of linkage mapping by allowing direct association of single markers to the quantitative trait, as in genome wide association (GWAS) or the exploitation of population structure as in haplotypes or IBD mapping.

Nonetheless, the identification of QTL and its application in LD-MAS still remain limited by several factors. Marker density, at least at this stage, is not yet dense enough to unravel QTL for traits with low heritability (which would benefit the most from MAS). Furthermore, a somehow simplistic mode of action for QTL is still assumed, which does not account for complex interactions among genes. Although several QTL studies have incorporated interacting QTL (Carlborg et al., 2000, 2003; Carlborg and Haley, 2004), the overall feasibility of these approaches is greatly limited by the growing complexity of models for interaction spanning more than a handful of genes. Furthermore, in any MAS selection scheme, markers explain only a portion of the total variance. This amount depends on the QTL effects and the number of markers included in the scheme, but for the vast majority of cases, a low number of markers with validated associations typically explain a small proportion of the genetic variance in a trait (Goddard and Hayes, 2009). Meuwissen et al. (2001) using simulated data proposed a different approach, which is now commonly identified as genomic selection. In their approach, (hundred) thousands of SNP markers are utilized contemporarily assuming that all QTL are in LD with at least one marker. For each of the markers, an estimated effect is calculated, and prediction of the estimate of an individual genetic value are obtained as the sum of individual marker effects according to the formula

$$GEBV = \sum^n x_i \hat{g}_i$$

[24]

where $GEBV$ represents the genomic breeding value of an individual obtained from the simultaneous prediction of $n$ individual markers, $x_i$ represents the genotype of the individual at a particular marker (e.g. 0, 1, 2 copies of a specific allele); and $\hat{g}_i$ is the predicted SNP effect. The use of molecular data within this framework offers several advantages. Contrarily to MAS schemes, genomic selection is able to explain all the genetic variance regardless of population structure and power of association; the same is true for whole genome association. In addition, by exploiting the LD between markers and QTL, genomic selection does not need a particular family structure and can be utilized in a population or even, given a set of markers dense enough, across populations (de Roos et al., 2009). Selection based on relatives relies on the historical records of all individuals’ performances and relationships within a population. This in turn means that for traits measured late in life or sex-limited traits, generation interval is prolonged. Genomic selection offers the opportunity to rely on reference populations in which individuals have both genotypic and phenotypic information and to employ prediction equations developed within the reference population to screen selection candidates before any of their progeny (or themselves) has any production. This allows a considerable shortening of the generation interval and consequently, higher genetic progress (Schaeffer, 2006). An example of the advantage in reduction of generation interval for genomic selection in a dairy progeny testing scheme is reported in Figure 1.

While in traditional progeny test schemes (in black in the picture) pre-screening of individuals entering the progeny test is based on parental average (PA) and the overall generation interval is affected by the time necessary to collect information on the daughters of the sire, genomic selection (in red) capitalizes on prediction equations developed in a reference population, allowing a faster turnover and a reduction in generation interval of approximately 3 years.

Finally, genomic selection offers an advantage in terms of accuracy of prediction over traditional selection schemes, with an average increase of the genomic breeding value reliability over parent average (PA) of about 30% (http://aipl.arsusda.gov/eval/summary/comparexml_menu.cfm).

The use of SNP genotypes in whole genome-enabled selection programs has received notable attention recently. Considerable effort has been devoted to the genotyping of a large proportion of the North American dairy sire population and genomic evaluations are available for this population since Spring 2008. Similar efforts are underway in several European populations. This provides, as mentioned, accurate estimates of genomic breeding values for sires, at least for traditional traits, as well as considerable shortening of the generation interval.

The success of this approach, however, relies on highly accurate conventional breeding values based on hundreds or thousands of progeny from sires that have been genotyped. Therefore, at least initially, whole genome selection will be limited to traits that are routinely collected in the commercial cow population and several challenges remain to be overcome in the application of genomic selection concerning shape and size of the populations employed as training generations. Traits that are not routinely collected on a commercial population will require the assembly of

![Figure 1. Reduction in generation interval through genomic selection in dairy progeny testing schemes.](http://aipl.arsusda.gov/eval/summary/comparexml_menu.cfm)
extremely large reference populations of the order of several thousands of individuals in order to provide reliable estimates of marker effects (Goddard and Hayes, 2009). Since population structure plays a role in the association of different markers with QTL, the transfer of prediction equations across populations is not straightforward (de Roos et al., 2009). Moreover, linkage disequilibrium between markers and QTL decreases over time and marker effects will need to be re-estimated over time (Muir, 2007).

Other challenges in the application of genomic selection are related to the modeling approach employed. It is envisaged that within a couple of years, the genomic approach will cover all populations and selection schemes, and the competition will rely, as in the past for the phenotypic selection, on developing the “best” genomic EBV estimation and genomic selection scheme. So far, genomic selection has been developed in practice only on dairy populations. In these populations, a two tier approach is employed in which information from traditionally calculated breeding values for sires under the form of daughter yield deviations or de-regressed proofs is employed to obtain estimates of genomic breeding values (VanRaden, 2008; VanRaden et al., 2009). Two large model classes are available in these settings. A traditional BLUP approach, in which DYD “meta-phenotypes” are included in the model and markers are incorporated in the model assuming equal variance among them according to the following simplified model

$$y = \mu + \sum_{i=1}^{p} X g_i + e$$  

[25]

where $E(g_i) = N(0, \sigma^2_i)$. The model is essentially an extension of Fisher’s model to a finite number of loci with the support of molecular data. A second Bayesian approach still makes uses of this two tier structure, but relies on the knowledge of the prior distribution of QTL effects, for which several markers will contribute negligibly to the overall variance. By utilizing appropriate priors for the marker variances, trivial effects can then be shrunk toward 0. This second approach is extremely appealing since it intrinsically makes use of the biological knowledge we possess for some of these traits and somewhat represents a liaison between association mapping and MAS efforts. Although the overall architecture of these models remain similar, several different implementations have been developed differing in the shape of the priors to be employed (Meuwissen et al., 2001; Cleveland and Deeb, 2009; de los Campos et al., 2009) and the overall number of markers to be effectively employed (Habier et al., 2009).

The two-step approach, whether within the BLUP or Bayesian framework, remains an ad hoc and somewhat cumbersome solution. Recently, various authors (Legarra et al., 2009; Misztal et al., 2009) proposed a unified single step approach, which will automatically blend genomic and phenotypic information into a single set of equations. This should notably simplify the application of genomic selection, granting simultaneously better statistical properties, improved accuracies for non-genotyped animals, greater resistance to genomic selection bias, and the possibility to apply genomic selection in multiple traits and with a large variety of models (Misztal et al., 2009).

All the models proposed so far for whole genome selection (i.e., Meuwissen et al., 2001; Xu, 2003) are able to accommodate only additive and potentially dominance effects. Recently, Gianola and Van Kaam and Gonzalez-Recio et al. (Gianola and van Kaam, 2008; Gonzalez-Recio et al., 2008) have proposed a non-parametric method that may accommodate the complexity of multiple interacting QTL in whole genome selection, but large-scale application of that method has not yet occurred.

Another unresolved issue relates to the reliability calculation of GEBV. The reliability of genomic predictions can be defined as their squared correlation with the true genetic merit and indicates the proportion of the genetic variance that is explained. Currently reliability of GEBV is obtained using cross validation either from prediction error variance (Su et al., 2010) or by using older bulls to predict young ones (VanRaden et al., 2009b). Alternatively, reliabilities can be obtained by inverting the left-hand sides of the mixed model equations. Furthermore, as reliability relies heavily on the number of phenotypes, combining data sets from multiple populations may be attractive as a way to increase reliabilities, particularly when phenotypes are scarce. However, this strategy may also decrease reliabilities if the marker effects are very different between the populations (de Roos et al., 2009). Calus et al., (2010) investigated various methods of assessing GEBV reliabilities and research in this field is very active.

Expansion of genomic selection for international genetic evaluations has also recently been proposed with the use of GMACE procedures by VanRaden and Sullivan (2010) or by different consortia efforts, such as EUROGENOMICS (EU-Holstein) or INTERGENOMICS (World Brown Association).

Conclusions

Genomic selection revolutionized the genetic improvement programs according to a new paradigm: from phenotype to predict genetic value (phenotypic selection), to genotype to predict expected performance (genomic selection), and is now a reality implemented in several dairy populations across the world. Genomic selection promises to overcome the bottlenecks of large routine data collection for traits difficult or expensive to measure, such as those related to food safety and quality, animal health, and responses to climate changes. Quality and homogeneity of data collected will play an important role in the ability of pooling and validating results from different populations. Several aspects of genomic selection still need to be clarified. The management of potentially increased inbreeding, the incorporation of genomic evaluation in total merit indexes, the evaluation of selective genotyping strategies with the availability of higher and lower density marker panels, and the application of SNP predictions for precision matchings represent only a few of the challenges that researchers will face in the near future. New analysis methods along with the refinement of the existing ones will be required to unravel the elusive nature of complex traits. For how daunting these tasks may seem, the collective knowledge this field has gained in such a short time period represents a formidable starting point.

References

Baker, R.J., 1986. Selection Indices in Plant Breeding. CRC Press, Boca Raton, FL, USA.
Brotherstone, S., Hill, W.G., 1986. Heterogeneity of variance among herds for milk production. Anim. Prod. 42:297-303.
Calus, M.P.L., Mulder, H.A., Verbly, K., Veerkamp, R.F., 2010. Estimating reliabilities of genomic breeding values. pp 198-201 in Proc. Interbull Int. Workshop Genomic Information in Genetic Evaluations, Upssala, Sweden.
Carlborg, O., Andersson, L., Kinghorn, B., 2000. The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. Genetics 155:2003-2010.
Carlborg, O., Haley, C.S., 2004. Epistasis: too often neglected in complex trait studies? Nat. Rev. Genet. 5:618-625.
Carlborg, O., Kerje, S., Schutz, K., Jacobsson,
New challenges in dairy genetics

b gster selection using low-density marker panels. Genetics 182:343-353.

Havill, D.A., 1975. Index selection with proportionality constraints. Biometrics 31:223-225.

Havill, D.A., Mee, R.W., 1984. A mixed-model procedure for analyzing ordered categorical data. Biometrics 40:393-408.

Hazel, L., 1943. The genetic basis for constructing selection indexes. Genetics 28:476-490.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447.

Henderson, C.R., 1984. Application of Linear Models in Animal Breeding. University of Guelph Press, Ontario, Canada.

Hill, W.G., 1984. On selection among groups with heterogeneous variance. Anim. Prod. 39:473-477.

James, J.W., 1968. Index selection with restrictions. Biometrics 24:1015-1018.

Jamrozik, J., Kistemaker, G.J., Dekkers, J.C., Schaeffer, L.R., 1997. Comparison of possible covariates for use in a random regression model for analyses of test day yields. J. Dairy Sci. 80:2550-2556.

Kennedy, B.W., 1989. Use of mixed model methodology in analysis of designed experiments. In: D. Gianola and K. Hammond (eds.) Advances in Statistical Methods for the Genetic Improvement of Livestock. Springer, Berlin, Germany, pp 77-94.

Lande, R., Thompson, R., 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

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

Lin, C., 1978. Index selection for genetic improvement of quantitative characters. Theor. Appl. Genet. 52:49-56.

Lynch, M., Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, MA, USA.

Mee, R.W., Havill, D.A., 1982. Maximum-likelihood estimation for an ordered categorical response model. Biometrics 38:1115 (abstr.).

Meuwissen, T.H.E., Goddard, M.E., 2000. Fine mapping of quantitative trait loci using linkage disequilibrium with closely linked marker loci. Genet. 155:421-430.

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

L., Jensen, P., Andersson, L., 2003. A global search reveals epistatic interaction between QTL for early growth in the chicken. Genome Res. 13:413-421.

Cassandro, M., Canavesi, F., Brandts, A., Carnier, P., Gallo, L., Bittante, G., Baghino, A., 1996. Trend of within country sire variance and potential impact on International evaluations for production traits. Interbull Meet. Proc. Bulletin 14:12-17.

Cassandro, M., Miglior, F., Carnier, P., Bittante, G., Canavesi, F., Santus, E., Banos, G., 1997. Effect of standardisation of within country-year sire variance of de-regressed proofs on international evaluations. Interbull Meet. Proc. Bulletin 16:12-20.

Cleveland, M.A., Deeb, N., 2009. Evaluation of a genome-wide approach to multiple marker association considering different marker densities. BMC Proc. 3 (Suppl 1):S3.

Cumberland, G.D., Riddick, L., Vinson, R., 1987. Earlobe cresces and coronary atherosclerosis. The view from forensic pathology. Am. J. Forensic Med. Pathol. 8:9-11.

Cunningham, E.P., 1969. The relative efficiencies of selection Indexes. Acta. Agr. S cand. 15:45-48.

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

de Roos, A.P., Hayes, B.J., Goddard, M.E., 2009. Reliability of genomic predictions across multiple populations. Genetics 183:1545-1553.

Dekkers, J.C., 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82:E313-E328.

Dekkers, J.C., Hospital, F., 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3:22-32.

Djemali, M., Berger, P.J., Freeman, A., 1987. Ordered categorical sire evaluation for dystocia in Holsteins. J. Dairy Sci. 70:2374-2384.

Durocq, V., Quaas, R.L., Pollock, E.J., Casella, G., 1988a. Length of productive life of dairy cows: I. Justification of a Weibull model. J. Dairy Sci. 71:3061-3070.

Durocq, V., Quaas, R.L., Pollock, E.J., Casella, G., 1988b. Length of productive life of dairy cows. 2. Variance component estimation and sire evaluation. J. Dairy Sci. 71:3071-3079.

Elston, R.C., 1963. A weight-free index for purpose of ranking or selection with respect to several traits at a time. Biometrics 19:85-97.

Fernando, R., Grossman, M., 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467-477.

Fiske, W.F., 2004. Comparison of performance records and national breeding values as input into international genetic evaluation. J. Dairy Sci. 87:2709-2719.

Fisher, R.A., 1918. The correlation between relatives on the supposition of mendelian inheritance. T. R. Soc. Edinb. 52:399-433.

Foulley, J.L., Gianola, D., 1984. Estimation of genetic merit from bivariate all or none responses. Genet. Sel. Evol. 16:285-306.

Foulley, J.L., Gianola, D., Thompson, R., 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth-weight and pelvic opening. Genet. Sel. Evol. 15:401-423.

Gelfand, A.E., Hills, S.E., Racinepoone, A., Smith, A.F.M., 1990. Illustration of Bayesian-inference in normal data models using Gibbs Sampling. J. Am. Stat. Assoc. 85:972-985.

Gianola, D., Fernando, R.L., 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63:217-244.

Gianola, D., Foulley, J.L., 1983. Sire evaluation for ordered categorical-data with a threshold model. Genet. Sel. Evol. 15:201-223.

Gianola, D., Foulley, J.L., Fernando, R.L., Henderson, C.R., Weigel, K.A., 1992. Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. J. Dairy Sci. 75:2805-2823.

Gianola, D., van Kaam, J.B.C.H.M., 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178:2305-2313.

Goddard, M.E., Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381-391.

Gonzalez-Recio, O., Alenda, R., Chang, Y.M., González-Recio, O., Gianola, D., Long, N., Weigel, K.A., Rosa, G.J.M., Avendano, S., 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. Genetics 178:2305-2313.
Miglior, F., Weigel, K., Banos, G., 1997. Impact of heterogeneous variance on dairy cattle breeding schemes: A simulation study. Livest. Prod. Sci. 33:31-41.

Misztal, I., Legarra, A., Aguilar, I., 2009. Use of test-day indices and computational methods for different test-day evaluation methods. J. Dairy Sci. 92:4648-4655.

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

Patak, E., Schaeffer, L.R., 1993. Use of test-day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34:23-34.

Quaas, R.L., Pollak, E.J., 1980. Mixed model methodology for farm and ranch beef cattle testing programs. J. Anim. Sci. 51:1277-1287.

Rendel, J.M., 1954. The use of regressions to increase heritability. Aust. J. Biol. Sci. 7:368-376.

Rönningen, K., Van Vleck, L.D., 1985. Selection index theory with practical applications. In: A.B. Chapman (ed.) World Animal Science, Vol A4. Elsevier, Amsterdam, The Netherlands, pp 187-222.

Sales, J., Hill, W.G., 1976. Effect of sampling errors on efficiency of selection indexes. 2. Use of information on associated traits for improvement of a single important trait. Anim. Prod. 23:1-14.

Sax, K., 1923. The association of size differences with seed-coat pattern and pigmentation in Phaseolus Vulgaris. Genetics 8:552-560.

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

Searle, S.R., 1965. The value of indirect selection: I. Mass selection. Biometrics 21:682-707.

Shook, G.E., 2006. Major advances in determining appropriate selection goals. J. Dairy Sci. 89:1349-1361.

Smith, H.F., 1936. A discriminant function for plant selection. Ann. Eugenic. 7:240-250.

Su, G., Guldbrautensen, B., Gregersen, V.R., Lund, M.S., 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J. Dairy Sci. 93:1175-1183.

Swalve, H.H., 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. J. Dairy Sci. 83:1115-1124.

Tallis, G.M., 1962. A selection index for optimum genotype. Biometrics 18:120-122.

Tempe1man, R.J., 1998. Generalized linear mixed models in dairy cattle breeding. J. Dairy Sci. 81:1428-1444.

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

VanRaden, P.M., Sullivan, P.G., 2010. International genomic evaluation methods for dairy cattle. Genet. Sel. Evol. 42:47.

VanRaden, P.M., Van Tassell, C.P., Wiggins, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., Spenkel, F.S., 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16-24.

Van Vleck, L.D., 1993. Selection Index and Improvement Program for North American Holstein bulls. J. Dairy Sci. 76:1680-1688.

Wright, S., 1921a. Systems of mating. I. The biometric relations between parent and offspring. Genetics 6:111-123.

Wright, S., 1921b. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. Genetics 6:124-143.

Wright, S., 1921c. Systems of mating. III. Assortative mating based on somatic resemblance. Genetics 6:144-161.

Wright, S., 1921d. Systems of mating. IV. The effects of selection. Genetics 6:162-166.

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

Yamada, Y., Yokouchi, K., Nishida, A., 1975. Selection index when genetic gains of individual traits are of primary concern. Jpn. J. Genet. 50:33-41.
