# Genomic selection: A paradigm shift in animal breeding

**Theo Meuwissen,* Ben Hayes,† and Mike Goddard‡**

*Norwegian University of Life Sciences, Ås, Norway

†Department of Economic Development, Jobs, Transport and Resources and Dairy Futures Cooperative Research Centre, Agribio, 5 Ring Road, Bundoora, VIC 3083, Australia; School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia

‡Department of Economic Development, Jobs, Transport and Resources and Dairy Futures Cooperative Research Centre, Agribio, 5 Ring Road, Bundoora, VIC 3083, Australia; Faculty of veterinary and agricultural sciences, University of Melbourne, Parkville, Australia

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022

## Implications

- Traditional marker-assisted selection (**MAS**) did not result in a widespread use of DNA information in animal breeding. The main reason was that the traits of interest in livestock production were much more complex than expected: they were determined by thousands of genes with small effects on phenotype. These effects were usually too small to be statistically significant and so were ignored.
- Genomic selection (**GS**) assumes that all markers might be linked to a gene affecting the trait and concentrates on estimating their effect rather than testing its significance. Three technological breakthroughs resulted in the current wide-spread use of DNA information in animal breeding: the development of the genomic selection technology, the discovery of massive numbers of genetic markers (single nucleotide polymorphisms; **SNPs**), and high-throughput technology to genotype animals for (hundreds of) thousands of SNPs in a cost-effective manner.
- Here we review current methods for GS, including how they deal with practical data, where genotypes are missing on a large scale. The use of whole-genome sequence data is anticipated, and its advantages and disadvantages are depicted. Current and predicted future impacts of GS on dairy and beef cattle, pigs, and poultry breeding are described. Finally, future directions for GS are discussed.
- It is anticipated that future GS applications will either be: within breed (**wbGS**), where accuracy is obtained by maintaining huge within-breed reference populations; or across breed (**abGS**) where accuracy is obtained from across-breed reference populations and high-density GS methods that focus on causative genomic regions. We argue that future GS applications will increasingly turn toward abGS.

**Key words:** cattle breeding; genetic improvement; genomic prediction; pig breeding; poultry breeding; whole-genome sequence

## Background

Animal breeding, i.e., the selective breeding for economically important traits, was traditionally based on phenotypic recordings. Best linear unbiased prediction (**BLUP**) combined individual records and those of relatives into estimates of breeding values (**EBV**). From 1990 onward, advances in molecular genetics held the promise that information at the DNA level would lead to more genetic improvement than using only phenotypic records. This resulted in research into MAS, which consists of two steps: 1) detect and (fine) map genes underlying the traits of interest, i.e., so called quantitative trait loci (**QTL**); 2) include the QTL information into the BLUP-EBV (Fernando and Grossman,1989).

The QTL mapping step (1) was successful in the sense that most mapping studies detected QTL. But the repeatability of the mapping studies was low, i.e., QTL positions moved/(dis)appeared from one study to the next. One reason for this is that the majority of QTL have very small effects. When this is combined with testing a large number of markers, there is a marked "Beavis effect" in which the estimated effect of significant markers is overestimated (Beavis, 1994). For instance, if we test 100 markers for their statistical significance using a *P*-value of 1%, we expect one (false) positive result even if all true marker effects are zero. Conversely, if all of the markers have very small effects, few (randomly picked) markers will reach higher levels of significance and most will fail to reach the threshold and be declared nonsignificant. In genome-wide association studies (**GWAS**), the number of tests equals the number of genotyped independent SNPs, which is typically many thousands in livestock and hundreds of thousands in human genetics. With so many SNPs, the multiple-testing problem becomes so large that in human genetics, *P*-values of $< 5 \times 10^{-8}$ are commonly used. In addition, human genetics journals demand a confirmation of the QTL in an independent dataset.

These very stringent tests resulted in only the largest QTL being found. For some traits, such large QTL were detected, e.g., DGAT1 affecting fat content in milk (Grisart et al., 2001) and CDH1 affecting infectious pancreatic necrosis virus (**IPNV**) resistance in Atlantic salmon (Moen et al., 2015). However, for many other traits, no reliable QTL were found, and less than 10% of the variation of the overall breeding objective, i.e., a combination of all the economically important traits, was explained by QTL. This was even the case for dairy cattle, where many powerful QTL mapping studies were conducted. Less than 10% of the genetic variance of the breeding objective explained by QTL implied that more than 90% of the genetic differences between animals had to be handled by traditional selection. Hence, by 2005, the uptake of MAS in livestock breeding was very limited. In human genetics, the result that very powerful GWAS studies (e.g., 160,000 individuals genotyped for 500,000 SNPs) explained only a (very) limited fraction of the total genetic variance was termed the missing heritability paradox (Manolio et al., 2009), i.e., a large part

© Meuwissen, Hayes, and Goddard.
doi:10.2527/af.2016-0002

of the heritable variation was not accounted for by these powerful GWAS studies.

Many explanations for the missing heritability problem have been published (Manolio et al., 2009). The most likely explanation seems to be a combination of very stringent statistical tests and many genes with small effects affecting the traits, i.e., the gene effects are too small to pass the stringent statistical tests despite the large number of genotyped individuals. All these small genes together explain the vast majority of the genetic variation for most traits (Yang et al., 2010). In 2001, Hayes and Goddard (2001) predicted 50–100 genes affected dairy traits, which was considered a high estimate at that time. Based on current GWAS and genomic selection results, we believe that dairy traits are affected by 2,000–10,000 genes. Thus, the number of genes that are believed to affect complex traits, such as dairy traits, has increased ~100 fold during the last 15 yr, i.e., complex traits turned out to be much more complex than expected one to two decades ago. Many genes affecting a trait implies that individual genes have small effects, which limits the efficiency of the MAS approach.


Three breakthroughs have resulted in the current widespread use of DNA information: 1) the GS methodology (Meuwissen et al., 2001), 2) the identification of many thousands of SNP markers, and 3) SNP-chip genotyping technologies that render the genotyping of all these SNPs cost effective. In MAS, a small number of significant markers were used, and the rest were treated as having zero effect. In GS, the effects of all these SNPs are estimated simultaneously without any significance testing. If there are ~10,000 genes affecting a trait, there are genes everywhere on the genome, which may be associated with many thousands of SNPs distributed across the genome. Hence, the assumption that all SNPs have an effect may be approximately valid, and we should change our focus from significance testing to estimating the effects of all markers. The (second generation) sequencing efforts that have resulted in discovery of the genome sequence of many of the livestock species have as a by-product revealed many thousands of SNP markers. In cattle, the 1,000-bulls sequencing project has revealed 30+ million SNP markers (Daetwyler et al., 2014). The SNP-chip genotyping technologies were mainly developed by Illumina and Affymetrix, and first-generation SNP chips contained typically ~50,000 SNPs for most livestock species.

In GS, a reference population is genotyped and recorded for the trait to estimate SNP effects. Next, selection candidates are genotyped, and by combining their genotypes with the estimated effects, genomic EBV (**GEBV**) are estimated for the selection candidates. It may be noted, that the GS approach does not require pedigree recording, which was essential to traditional BLUP-EBV, and that the elite breeding animals, i.e., the selection candidates, are not necessarily trait recorded. In traditional breeding, the elite breeding animals were as accurately as possible trait and pedigree recorded. This potential to decouple accurate recording from the elite breeding population makes it possible to completely redesign the breeding scheme, and consequently GS has resulted in a paradigm shift in animal breeding. Our goal here is to describe the GS method in more detail for a general scientific (non-geneticists) audience. In addition, we will describe current and predict future impacts of GS on dairy and beef cattle, pigs, and poultry breeding.

## Genomic Selection Methods

All SNP effects are simultaneously estimated in a reference population, which is genotyped and phenotyped using the statistical model (assuming 50,000 SNPs):

$$y_i = \boldsymbol{\mu} + X_{1i} \times b_1 + X_{2i} \times b_2 + \ldots. + X_{50000i} \times b_{50000} + e_i$$

where $y_i$ is phenotype of animal i; μ is the overall mean; $X_{1i}$ is the genotype of animal i for marker 1; and $e_i$ is the residual. Since usually we have < 50,000 reference animals, we cannot estimate 50,000 SNP effects if they are treated as fixed effects, i.e., using traditional statistical methods. This problem is solved in GS by treating the SNP effects as random effects drawn from a known distribution. This can be viewed as a Bayesian approach, where prior information on the SNP effects is added to make all effects estimable. A commonly used prior assumption is that SNP effects are normally distributed with mean 0 and a constant variance (which is the total genetic variance divided by 50,000). In effect, this method uses BLUP to estimate SNP effects, and the method is sometimes called SNP-BLUP. The genomic breeding value of selection candidate j is predicted as:

$$GEBV_j = X_{1j} \times {}_1 + X_{2j} \times {}_2 + \ldots. + X_{50000j} \times {}_{50000}$$

where $_1$ is the estimate of the effect of SNP 1; and $X_{1j}$ is the genotype of animal j for SNP 1.

## The GBLUP Method

In traditional BLUP, EBV are estimated using phenotypes and family relationships, which are based on the pedigree of the animals. In GBLUP,

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022

GEBV are estimated using phenotypes and genomic relationships, which are based on genome-wide dense marker data. The genomic relationship between animals 1 and 2 is calculated as the correlation between their SNP genotypes $X_{j1}$ and $X_{j2}$ across all the SNPs j. The GBLUP method is thus very similar to traditional BLUP, except that pedigree relationships are replaced by genomic relationships. A practical advantage of the GBLUP approach is that all the traditional BLUP methods and software can still be applied: we only need to replace pedigree by genomic relationships.

The pedigree relationship between two fullsibs is 0.5, which means that two full sibs are expected to have 50% of their alleles in common. However, in real life, two fullsibs may share 60% of their alleles or 40%, and this deviation from the pedigree-based expectation of 50% will be detected by dense marker genotyping. Thus, GBLUP is more accurate than traditional BLUP because genomic relationships are more accurate than pedigree-based relationships. The latter requires genomic relationship estimates to be based on a sufficiently large number of SNPs. For livestock and relationships within a breed, 50,000 SNPs distributed across the entire genome seems to suffice (Goddard et al., 2011). Relationships across breeds are small and require a larger number of SNPs to be used.

The statistical model for the GBLUP method is:

$$y_i = \boldsymbol{\mu} + u_i + e_i$$

where $u_i$ is the breeding value of animal i. The GBLUP and SNP-BLUP breeding values are equivalent if we define $u_i$ as:

$$u_i = X_{1i} \times b_1 + X_{2i} \times b_2 + \ldots. + X_{50000i} \times b_{50000}$$

This definition of $u_i$ has consequences for the covariance between two animals $u_i$ and $u_j$, which becomes:

$$(X_{1i} \times X_{1j} + X_{2i} \times X_{2j} + \ldots. + X_{50000i} \times X_{50000j}) s_b^2 =$$
$$(X_{1i} \times X_{1j} + X_{2i} \times X_{2j} + \ldots. + X_{50000i} \times X_{50000j})/50000$$

## Glossary

| | |
|---|---|
| **BLUP** | best linear unbiased prediction |
| **GBLUP** | genomic best linear unbiased prediction |
| **ssGBLUP** | single-step genomic best linear unbiased prediction |
| **SNP-BLUP** | best linear unbiased prediction of SNP effects |
| **EBV** | estimated breeding value |
| **GEBV** | genomic estimated breeding value |
| **GS** | genomic selection |
| **abGS** | across-breed genomic selection |
| **wbGS** | within-breed genomic selection |
| **APY** | Ancestor, Proven, Young Bull algorithm |
| **GWAS** | genome-wide association study |
| **LD** | linkage disequilibrium |
| **MAS** | marker-assisted selection |
| **MOET** | multiple ovulation and embryo transfer |
| **QTL** | quantitative trait locus |
| **SNP** | single nucleotide polymorphism |
| **WGS** | whole-genome sequence |

where the total genetic variance is assumed 1 (for simplicity) and the variance per SNP is then $s_b^2 = 1/50000$. If we standardize the genotypes $X_{ki}$ such that they have mean 0 and standard deviation 1 within every SNP k, the above formula calculates the correlation between SNP genotypes, i.e., the genomic relationship between the animals i and j. (The usual corrections for means and standard deviations in the correlation coefficient calculation are not needed here because the SNP genotypes are scaled so that their mean is 0 and standard deviation is 1.) When parametrized in this way, the SNP-BLUP and GBLUP model imply the same covariances between animals, and thus also identical regression coefficients of the records on the genetic value of animals. The latter implies that, when parameters are carefully adjusted, SNP-BLUP and GBLUP yield identical GEBV, i.e., the methods are said to be equivalent. More formal derivations of the equivalence of GBLUP and SNP-BLUP can be found in the literature (Habier et al., 2007; VanRaden, 2008; Goddard, 2009).

The computational requirements of GBLUP and SNP-BLUP may be very different. SNP-BLUP requires the estimation of 50,000 SNP effects, and thus the solving of a set of 50,000 equations, whereas GBLUP requires the estimation of N GEBV and solving of N equations, where N is the number of animals. Since usually the number of genotyped animals is smaller than 50,000, the GBLUP method is (computationally) preferred. In the future, the number of genotyped animals is expected to increase dramatically, so it may well be that the SNP-BLUP method becomes the method of choice. However, other, non-BLUP methods, may also gain popularity as shown in the following sections.

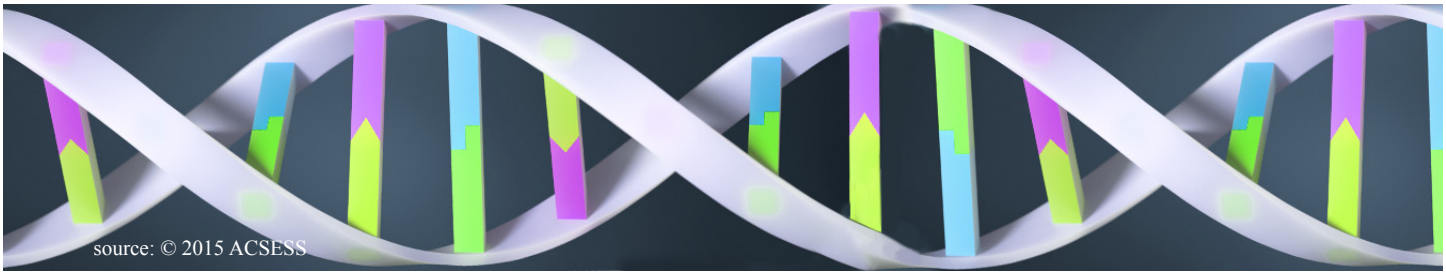
## Nonlinear Methods for Genomic Selection

The prior information in SNP-BLUP (and implicitly GBLUP) assumes that SNP effects are normally distributed with the same variance for every SNP. This assumption leads to BLUP estimates for the SNP effects that are a linear combination of all the observed phenotypes. Biologically, we may expect that some SNPs, that are close to a gene, have an effect and many others have no effect. A number of methods have been developed that incorporate prior information that assumes that a fraction $\pi$ of the SNPs have an effect and a fraction $(1-\pi)$ have no effect at all. The model used for these methods is:

$$y_i = \boldsymbol{\mu} + I_1 \times X_{1i} \times b_1 + I_2 \times X_{2i} \times b_2 + \ldots.$$
$$+ I_{50000} \times X_{50000i} \times b_{50000} + e_i$$

where $I_j$ is an indicator variable with values 0 or 1 indicating whether SNP j is having an effect or not. BayesC assumes that the SNPs with effects are normally distributed with constant variance (Habier et al., 2011) and is thus closest to SNP-BLUP. BayesB uses the t-distribution as prior for the SNPs with effects, which allows for some SNPs to have very big effects (Meuwissen et al., 2001). BayesR assumes a mixture of normal distributions for the effective SNPs, which also allows for some SNPs with very big effects, namely those that are drawn from the distribution with largest variance (Erbe et al., 2012). The estimated SNP effects from these methods are no longer a linear combination of phenotypes. Other nonlinear estimation methods are BayesA (Meuwissen et al., 2001), the LASSO, Bayesian Lasso, and the elastic net (Hastie et al., 2009). (The nonlinear methods are sometimes called "Bayesian" methods because they use a prior distribution of SNP effects, but SNP-BLUP also uses a prior distribution, which is assumed to be a normal distribution.)

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022




The prior distribution of SNP effects used by the nonlinear methods makes a lot more sense biologically than assuming that all SNPs have an effect and that all effects are very small. In computer simulation studies, the nonlinear methods clearly outperform GBLUP (Meuwissen and Goddard, 2010), but in real data, nonlinear methods are somewhat superior for some traits but not all (Erbe et al., 2012). This may be explained by the following: 1) there are many genes affecting the economically important traits, so that assuming all SNPs are having an effect is approximately true; 2) linkage disequilibrium (the non-random association between two loci) extends over large genomic distances in livestock populations, such that many SNPs are associated with a gene; and 3) the SNP density is not high enough, so that each QTL can be explained by a single SNP and so many SNPs are needed to jointly explain the QTL effect. Consequently, the superiority of nonlinear methods over GBLUP becomes clearer when the 50k SNP chip is replaced by a high-density 777k SNP chip and when the data includes multiple breeds (Brøndum et al., 2015). The combination of explanations (1) and (2), i.e., there are many genes relative to the extent of the linkage disequilibrium, mainly explains the good performance of GBLUP within breeds.

## Sequence Data

Genomic selection based on SNP chip genotypes relies on linkage disequilibrium (**LD**) between the QTL and the SNPs, i.e., associations between SNPs and QTL. Increasing the density of SNPs increases the probability that any QTL has a SNP that is in perfect LD with it. The ultimate density is to replace SNP genotypes with whole-genome sequence (**WGS**) data. In the latter case, the causative mutations are expected to be present in the sequence data, and thus, GS can act on the causative mutations directly, instead of having to rely on LD between markers and causative mutations. However, these mutations are hidden among many millions of SNPs with no effect. It may thus be expected, and simulation studies have shown (Meuwissen and Goddard, 2010), that the nonlinear GS methods, which assume that many SNPs have no effect, yield substantially higher accuracies than GBLUP when using WGS data.

Recently, Brøndum et al. (2015) demonstrated small (2–5%) increases in the accuracy of GEBV with sequence data. That current WGS data do not result in substantial improvements in accuracies of GEBV may be explained as follows. First, the GBLUP method is expected to yield little improvement when going from 777k to WGS data since the genomic relationships are accurately estimated with 777k data and WGS will hardly improve the accuracy of the relationships and thus GEBV. However, the nonlinear GS methods attempt to identify the causal SNPs and are expected to benefit substantially from WGS data. Second, current WGS data are not very accurate, either due to imperfect genotype calling, the extensive reliance on SNP imputation (see next section), or structural genomic variations, which are difficult to assess by short reads of sequences. The inaccuracies in the WGS data may compensate for the benefits of higher SNP density. Third, long-range LD may be extensive in the reference population animals, causing large chromosomal segments or haplotypes to be common. Consequently, there will be many combinations of SNPs that explain the effect of the haplotype as well as the causal mutations. Each statistical method will chose a combination of SNP effects that best fits its prior assumptions, but they may all give the same prediction of the haplotype effect. However, if the range of the LD is reduced, e.g., by using a reference population that is less closely related, the nonlinear methods that focus on the causal mutations may give greater accuracy than GBLUP, which uses all sequence variants equally. Another problem is that present-day computers struggle to store and handle these massive amounts of data, especially if WGS is to be collected on many animals. Despite current issues with the efficient use of WGS data, it is expected that WGS data will be the future's genotype data because, if the sequencing costs continue to fall, WGS may become the most effective genotyping method (Gorjanc et al., 2015).

## The Imputation of Missing Genotypes

After SNP-chip genotyping, some of the genotypes will be missing. This is solved by a process called genotype imputation. Based on the known genotypes of the animals, the haplotype that the animal carries is recognized since the same haplotype was also observed in other animals. Thus, the missing genotype can be read from the genotype of these other animals, which carry the same haplotype. Software for imputation includes Beagle (Browning and Browning, 2007), Fimpute (Sargolzaei et al., 2014), and Alphaimpute (Hickey et al., 2012).

Imputation methods can also be used in combination with sparse, but cheap, SNP chips. Key ancestors are genotyped with the dense, but expensive, chip to identify the haplotypes in the population. Next, large numbers of descendants are genotyped with a sparse, cheap SNP chip. The sparse chip has enough SNPs to recognize which of the haplotypes the animal carries. Since the haplotypes are known at high density, the missing genotypes can be imputed.

The same strategy is employed to obtain WGS data on many animals: the 1,000-bull-genome project (Daetwyler et al., 2014) collects a set of sequenced bulls across breeds, which is used to identify (hopefully all) bovine haplotypes and their sequences. Next, many animals are genotyped with SNP chips, the bovine haplotypes that they carry are recognized, and their WGS data are imputed. Another option is to sequence the descendants at low coverage. In this case, the low coverage sequence should be just enough to recognize the haplotypes (Gorjanc et al., 2015). The 1,000 bull genomes project demonstrated that accurate imputation of sequence genotypes was possible for SNPs (and other variants) with high minor allele frequency. For SNPs with low minor allele frequency however, accuracy of imputation was

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022

poor. Druet et al. (2014) demonstrated (in simulation) that if large numbers of ancestors were sequenced, at relatively low coverage (four- to sixfold), accuracy of imputing genotypes for these rare SNPs was improved. van Binsbergen et al. (2014) clearly demonstrated that imputing 50K genotypes first to 800K, then to sequence, resulted in higher accuracy of imputation than if 50K genotypes were imputed directly to sequence.

## Ungenotyped Animals

In genomic selection, many (probably most) animals are not genotyped, but we need to include their phenotypic information in the breeding value estimation. At least, traditional selection would use such information. One way to do this is by multiple-step GS: in step 1, pseudo-phenotypes are calculated for the genotyped animals where the pseudo-phenotype of animal i includes information (records) on its ungenotyped relatives; in step 2, genomic prediction is performed using the pseudo-records and their genotypes; and in step 3, the traditional EBV and GEBV are combined into a total EBV (e.g., VanRaden, 2008). As an example of a pseudo-record, the average production of the daughters of a bull can be used. Here, the bull is genotyped but not phenotyped whereas his daughters are phenotyped but not genotyped. Since the data are handled in multiple steps, this method is clearly suboptimal. However, in practice, good GS accuracies have been achieved using this method.

In single-step GBLUP (**ssGBLUP**), all data are accounted for in a single estimation step (see Legarra et al., 2014 for a review). When moving from traditional BLUP to GBLUP, we replace the entire matrix of pedigree relationships with genomic relationships (see above). An obvious idea is to replace pedigree with genomic relationships where available and retain the pedigree relationships where we do not have genomic relationships. However, if genotyping shows that e.g., some animals in different families are more related than expected based on pedigree, then other ungenotyped animals in these families are probably also more related than expected. The correct relationship matrix can be obtained by starting with the genotyped animals and then using the pedigree to calculate relationships involving ungenotyped descendants of these genotyped animals, i.e., going down the pedigree and accounting for the marker-based relationships of the ancestors of the pedigree. The same idea can also be used up the pedigree, i.e., when ancestors are non-genotyped although it is not optimal in this case (Meuwissen et al., 2011). In dairy cattle, ssGBLUP yields 0–2% more accuracy than multistep methods (Legarra et al., 2014), but for other species, which are less dominated by large sire families (i.e., where daughter averages are less able to summarize family information), the difference in accuracy between ssGBLUP and the multistep methods may be larger. A shortcoming of the single-step method is that it so far does not work for nonlinear estimation although some solutions to single-step nonlinear estimation have been proposed in the literature (Liu et al., 2014; Legarra and Ducrocq, 2012).

In most studies, increases in reliability due to single step, over a pure genomic model, are small (e.g., Koivula et al., 2012). A more important feature of single-step models may be that they can account for pre-selection of young genotyped bulls, which could otherwise cause bias in the GEBV (Vitezica et al., 2011). Until recently, the requirement that the **G** matrix must be inverted directly limited the size of the dataset to which ssBLUP could be applied. The Ancestor, Proven, Young Bull algorithm (**APY**) uses recursion to build a large component of the $\mathbf{G}^{-1}$ matrix directly, overcoming this limitation and expanding the application of ssBLUP to millions of genotyped animals (Fragomeni et al., 2015) but at the expense of some approximation in $G^{-1}$. For the future, there is a clear need for a single-step method that uses a nonlinear statistical method on sequence level data.

## Implementation of Genomic Selection in Livestock Industries

### *Genomic selection in dairy cattle*

The accuracy of genomic prediction in dairy cattle exceeds 0.8 for production traits and 0.7 for fertility, longevity, somatic cell count, and other traits (e.g., Wiggans et al., 2011; Lund et al., 2011). These high accuracies reflect the large reference populations for each breed that have been assembled to enable genomic predictions and the fact that many of the animals in the reference populations are progeny-tested bulls with highly accurate phenotypes from average daughter performance. In addition, the GEBV are often used to predict close relatives of animals in the reference population. A feature of dairy genomic predictions is collaboration between countries to assemble these large reference sets, with three consortiums established (Eurogenomics, including the Netherlands, Germany, France, the Nordic countries, Spain, and Poland; The North American Consortium including USA, Canada, Italy, and Great Britain; and a "rest of the world" consortium consisting of a number of remaining countries).

The high accuracies of genomic prediction and relatively low cost of obtaining the genomic predictions from low-density genotyping followed by imputation, has resulted in very large numbers of selection candidates being genotyped. Worldwide, approximately 2 million dairy cattle have now been genotyped for the purposes of genomic prediction. In the USA alone, 934,780 Holstein animals, 120,439 Jersey animals, 19,588 Brown Swiss, and 4,767 Aryshire animals have been genotyped (Wiggans, personal communication, https://www.cdcb.us/Genotype/cur_density.html). Similar numbers of animals have been genotyped by other countries combined, including 360,000 in France alone (Boichard, personal communication).

Implementing genomic selection in dairy cattle has resulted in increased genetic gain, which has now been demonstrated by genetic trend analysis in a number of countries. For example, in Canada, the rate of genetic gain has approximately doubled since genomic selection was introduced (VanDoormal, personal communication). There is also some suggestion that genomic selection has increased the rate of inbreeding per year (Schenkel, 2012). Maximizing genetic gain from genomic selection while constraining the rate of inbreeding will therefore be an important topic for future research.

Interestingly, the majority of the genotyped animals in many countries are now heifer calves. While genotyping young bull calves results in the greatest genetic gain, genotyping is now sufficiently cheap that genotyping heifer calves for the purposes of choosing which heifers to retain in the herd is profitable (Pryce and Hayes 2012; Weigel et al., 2012). The genotypes of the heifers can also be used when choosing bulls to which to mate them so that inbreeding of the resulting calf can be minimized.

When the selected heifers enter the herd and have herd recording data, they can be used in the reference population for genomic prediction (Wiggans et al., 2011). When the aim is to increase the size of the reference population to improve accuracy of genomic prediction, genotyping mature cows with good phenotypic records can help—Kemper et al. (2015) reported that adding 10,000 and 5,000 cows to reference sets used to evaluate Holstein and Jersey cattle respectively added a 5–8% increase in

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022

accuracy, depending on the trait. This relatively large increase probably reflects the smaller bull reference sets (4,000 and 1,000 for Holsteins and Jerseys, respectively) compared with some of the populations above.

## Genomic selection in beef cattle

In some beef breeds, genomic selection is now applied on a large scale. For example, in the USA, more than 52,000 Angus animals have now been genotyped for GEBV evaluation (Lourenco et al., 2015). In general, however, accuracies of genomic predictions in beef cattle have been lower than in dairy cattle. For instance, in their review, Van Eenennaam et al. (2014) reported accuracies in the range 0.3 to 0.7. The lower accuracy is because the reference populations are of higher quality in dairy cattle. In beef cattle, the reference population contains fewer animals within a breed, and these animals have not been progeny tested. In addition, the target population and validation animals may be less closely related to the reference population in beef cattle than in dairy cattle.

To compensate for the small number of reference animals within a breed, it is not uncommon to use a multi-breed reference population. Bolormaa et al. (2013b) found that this increased accuracy slightly (0.33 to 0.38) but not as much as if the same number of animals had been from the same breed. When Akanno et al. (2014) used a reference population of several pure breeds in USA to predict within a crossbred population in Canada, the accuracy was low. If the target breed is not included in the reference population, the accuracy is very low.

These disappointing results for prediction across breeds are not unexpected. De Roos et al. (2009) found that the phase of LD does not persist across breeds except at short genomic distances (e.g., < 10kb). Therefore, when using a 50k SNP panel, information from another breed is not expected to increase accuracy. Even if high density SNPs are used, information from another breed is much less useful than information from the target breed because animals of different breeds share much smaller chromosome segments than animals of the same breed. When BLUP is used to predict breeding values, the variance for a chromosome segment is assumed to be proportional to its length (or number of SNPs), and so small segments have lower variance and are estimated less accurately than larger segments. The situation is improved a little by using Bayesian methods that allow some SNPs (and therefore some segments) to have a larger effect than others. Then prediction can make better use of SNPs in high LD with the QTL, and this information may transfer across breeds (Bolormaa et al., 2013b; Khansefid et al., 2014).

The value of combining breeds in a reference population depends to some extent on QTL segregating in multiple breeds. Bolormaa et al. (2014) reported QTL in similar locations across a range of breeds, suggesting that QTL do segregate in multiple *Bos taurus* breeds. However, Bolormaa et al. (2013a) concluded that QTL seldom segregate in both *B. taurus* and *B. indicus*, although there are known exceptions such as PLAG1 and CAST [PLAG1 segregates in Australian Brahmans because it was introgressed from *B. taurus* (Fortes et al., 2013)].

Genomic selection has not been adopted as widely in beef as in dairy cattle breeding. This is partly because the accuracy is lower, but also because the economic advantages are not as great. Genomic selection is most advantageous for traits that are difficult to select for traditionally. It is less advantageous in beef than dairy because progeny testing is not needed for traits that can be measured on selection candidates at a young age such as growth rate. However, several important traits in beef cattle are difficult to select for such as feed conversion efficiency and beef quality. Because these traits are also expensive to record, it is costly to set up a large training population and there are no large companies that could justify this cost for their own breeding program. For these traits, a multi-breed training population and nonlinear analysis based on high-density SNPs or genome sequence data may be the best approach.

Despite these difficulties, genomic selection has been implemented in beef cattle. For instance, Angus EBVs in Australia and USA are calculated using DNA information if it is available. There are two ways in which this can be done. First, the genotypes can be provided to the organization that calculates EBVs who then calculate the prediction equation. Second, a commercial organization, such as Zoetis or GeneSeek, can provide the DNA testing service and use their own prediction equation to generate "marker breeding values" which are then transmitted to the genetic evaluation service for incorporation into EBVs. Both methods are in operation. The advantage of the first method is that the full dataset of phenotypes and genotypes can be used to derive the prediction equation and the DNA information can be fully integrated through one-step genomic prediction methods (e.g., ssGBLUP).

## Genomic selection in pig breeding

In pig breeding, the most important selection step is the selection of elite boars in the nucleus herd. (This may be at the boar test station in the case of cooperating pig breeders.) The boar test recordings come generally before the selection of the elite boars, so extra gains due to a reduction of the


source: © 2015 Stock.Adobe.com/fottoo

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022

generation interval are limited although still a ~25% reduction in generation interval may be realized by the introduction of GS (Bjarne Nielsen, personal communication, 2015). The implementation of GS in pig breeding is therefore mainly directed at traits whose recording is invasive such as slaughter quality, maternal traits that cannot be recorded on the boars, and crossbred performance, which cannot be recorded on the purebred animals.

With respect to the maternal traits, female sibs of the test boars are raised in nucleus herds, but their maternal trait recordings become available after the selection of the boars. However, GS for maternal traits can be based on aunts of the test boars. Male selection accuracies of ~50% can be achieved for the selection for maternal traits (Lillehammer et al., 2011). The selection for maternal traits competes with the selection for production traits such as growth rate and feed conversion efficiency, resulting in substantial increases in genetic gain for maternal traits accompanied with a somewhat reduced rate of gain for the production traits. Rates of gain for total merit increase moderately, but the direction of rate of gain complies much more closely to the direction as indicated by the breeding goal. The substantially increased progress for the maternal traits thus results in a more balanced and thus sustainable selection response.

With respect to slaughter traits, sibs of the test boars may be slaughtered and recorded for these traits before the test boars are selected. Thus, GS can be based on a reference population that is very close to the selection candidates, and thus high selection accuracies can be achieved. Also, here the extra gains will partly be at the expense of gains for the traditional production traits, but the direction of genetic change will comply much closer to the breeding goal, and thus can be sustained over a longer period into the future.

Pork is produced by crossbred pigs, but the elite breeding nucleus animals are selected for purebred performance in a favorable environment (e.g., a nucleus herd). The relationship between purebred production in very good environments and crossbred performance on less favorable environments varies between 0.4 to 0.7 (Esfandyari et al., 2015). This implies that only 40–70% of the genetic improvement realized in the nucleus will also result in improved performance in practice, e.g., if nucleus pigs grow 100 g/day faster due to genetic improvement, commercial pigs will only grow 40–70 g/day faster. By genotyping crossbred pigs and recording their performance in the commercial environment, GS can be used to improve purebred nucleus animals for crossbred performance under commercial circumstances. This requires across breed and crossbred genomic selection, which has not yet been demonstrated.

Hence, optimal across breed/purebred genomic selection methods need to be developed. Pig breeding companies are currently working toward such a solution for the direct genetic improvement of crossbred performance under practical conditions. The same approach can be used to select for traits that are relevant for an export market but are not recorded at the home market (e.g., resistance to some diseases). The approach will require an infrastructure where performance data and genotypes of practical animals are collected (across countries), transferred to the breeding value evaluation center, and used for selection in the nucleus.

### *Genomic selection in poultry*

In layers, there has actually been an experiment to test if genomic selection can achieve more rapid gains than traditional selection. In Wolc et al. (2015), a layer population was split into two sublines; one was submitted to conventional phenotypic selection, and one was selected based on genomic prediction. The experiment ran for 3 yr, in which time, four cycles of genomic selection and two of phenotypic selection were conducted. At the end of the 3-yr experiment, the two sublines were compared for multiple performance traits that are relevant for commercial egg production. The genomic selection line outperformed the phenotypic selection line for most of the 16 traits that were included in the index used for selection. Although the two programs were designed to achieve the same rate of inbreeding per year, Wolc et al. (2015) found that the realized inbreeding per year assessed from pedigree was higher in the genomic-selected line than in the conventionally selected line.

In broilers or meat poultry, the case for GS is not as obvious as in layers because most traits can be recorded on both sexes at an early age. However, the breeding companies are actively investigating the use of GS. Possible uses are for selection to improve crossbred performance in a commercial environment and for traits that cannot be recorded in the nucleus such as disease challenge tests.


source: © 2015 Stock.Adobe.com/buhanovskiy

## Future Directions

The cost of DNA testing is an impediment to its use in many cases. If this cost continues to fall, the use of DNA testing will expand. This will help to generate larger and more up-to-date reference populations. A problem will continue to exist for traits that are not routinely recorded. One-step evaluation methods are likely to become the norm. This will occur because they deliver more accurate EBVs and because cheaper DNA testing will lead to a higher proportion of the animals being tested.

Two methods could be used in the future to calculate genomic EBVs. The evaluation could be within breed (within breed GS; wbGS). In this case, medium- or low-density SNPs are enough and G- or SNP-BLUP can be used to calculate the prediction equation. Alternatively, the training population might consist of mul-

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022

tiple breeds and perhaps crosses (abGS). In this case, EBVs will be more accurate if dense SNPs are used and nonlinear methods are used to calculate the prediction equation. For wbGS, there seems to be little opportunity to improve the prediction other than getting larger reference populations. For abGS, there are many avenues for improvement. Genome sequence data can generate more accurate EBVs than dense SNPs because the causal variants are included in the data and we don't have to rely on LD. Increased biological knowledge about the effects of mutations at each site in the genome can be used to discover these causal variants.

If wbGS yields sufficiently accurate EBVs (e.g., > 0.9), there is no need to explore abGS. However, in the future, GS will be used for an increasing number of traits of which some will be difficult to record on a large scale (e.g., methane emissions), and this reduces the opportunities for large within-breed reference populations for wbGS. In the longer term, we thus believe that abGS will lead to more accurate EBVs for the overall breeding goal, which is stable over populations that vary in space and time. In case the future of GS is limited to wbGS, the number of breeds and lines within breeds will decrease because only the largest lines will have large enough training populations to generate accurate EBVs.

Genomic selection offers two opportunities, which have so far not been fully utilized. First, GS combined with reproductive technology could greatly decrease generation length, and in combination with multiple ovulation and embryo transfer (**MOET**), GS may be used to pick the best embryos to produce the next generation of animals (instead of random embryos). Second, we could train the prediction equation on commercial animals, rather than stud animals, which have been measured for the commercially relevant traits. For instance, commercial animals are often crossbreds, run under a harsher environment than the purebred stud animals. In addition, we can gather information on traits not measured at the stud level such as meat quality and disease resistance (e.g., if the outbreak of an infectious disease is a rare event). This implies a reduction of costs at the stud level due to less phenotypic and pedigree recording and an increase in costs to generate the reference dataset. This change may also be expected as the costs of genotyping and practical trait recording keep on falling (e.g., use of sensors in automatic milking systems). A commercial mechanism to fund this change is not yet apparent.

It seems likely that the GS paradigm shift in animal breeding will eventually lead to structural changes in the genetic improvement industry, but it may be too early to nominate what these changes might be. One possibility is that the number of businesses breeding cattle, sheep, and pigs will decrease as has already happened in poultry.

## Literature Cited


Akanno, E.C., G. Plastow, B.W. Woodward, S. Bauck, H. Okut, X.L. Wu et al. 2014. Reliability of molecular breeding values for Warner-Bratzler shear force and carcass traits of beef cattle—an independent validation study. J. Anim. Sci. 92(7):2896–2904. doi:10.2527/jas.2013-7374.

Bolormaa, S., J.E. Pryce, K.E. Kemper, B.J. Hayes, Y. Zhang, B. Tier et al. 2013a. Detection of quantitative trait loci in *Bos indicus* and *Bos taurus* cattle using genome-wide association studies. Genet. Sel. Evol. 45:43. doi:10.1186/1297-9686-45-43.

Bolormaa, S., J.E. Pryce, K. Kemper, K. Savin, B.J. Hayes, W. Barendse, et al. 2013b. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. J. Anim. Sci. 91(7):3088–3104. doi:10.2527/jas.2012-5827.

Bolormaa, S., J.E. Pryce, A. Reverter, Y. Zhang, W. Barendse, K. Kemper et al. 2014. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. PLoS Genet. 10(3):e1004198. doi:10.1371/journal.pgen.1004198.

Beavis, W.D. 1994. QTL analysis: Power, precision and accuracy. In: A.H. Paterson, editor, Molecular dissection of complex traits. CRC, Boca Raton, FL.

Brøndum, R.F., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, D. Boichard, and M.S. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J. Dairy Sci. 98(6):4107–4116.

Browning, S., and B. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097. doi:10.1086/521987.

Daetwyler, H.D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R.F. Brøndum, X. Liao et al. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46(8):858–865.

deRoos, A.P.W., B.J. Hayes, and M.E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics 183(4):1545–1553. doi:10.1534/genetics.109.104935.

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

Erbe, M., B.J. Hayes, L.K. Matukumalli, S. Goswami, P.J. Bowman, C.M. Reich et al. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95(7):4114–4129. doi:10.3168/jds.2011-5019.

Esfandyari, H., A.C. Sorensen, and P. Bijma. 2015. A crossbred reference population can improve the response to genomic selection for crossbred performance. Genet. Sel. Evol. 47:76. doi:10.1186/s12711-015-0155-z.

Fernando, R., and M. Grossman. 1989. Marker assisted selection using Best Linear Unbiased Prediction. Genet. Sel. Evol. 21:467–477. doi:10.1186/1297-9686-21-4-467.

Fortes, M.R.S., K. Kemper, S. Sasazaki, A. Reverter, J.E. Pryce, W. Barendse et al. 2013. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. Anim. Genet. 44(6):636–647. doi:10.1111/age.12075.

Fragomeni, B.O., D.A. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T.J. Lawlor, and I. Misztal. 2015. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. J. Dairy Sci. 98(6):4090–4094. doi:10.3168/jds.2014-9125.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica (The Hague) 136:245–257.

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

Gorjanc, G., M.A. Cleveland, R.D. Houston, and J.M. Hickey. 2015. Potential of genotyping-by-sequencing for genomic selection in livestock populations. Genet. Sel. Evol. 47:12. doi:10.1186/s12711-015-0102-z.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi et al. 2001. 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. doi:10.1101/gr.224202.

Habier, D, R.L. Fernando, and J.C.M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. 177(4):2389–2397. doi:DOI 10.1534/genetics.107.081190.

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

Hastie, T, R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics.

Hayes, B., and M.E. Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. 33(3):209–229. doi:10.1186/1297-9686-33-3-209.

Hickey, J., B. Kinghorn, B. Tier, J. van der Werf, and M. Cleveland. 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. Genet. Sel. Evol. 44(1):9. doi:10.1186/1297-9686-44-9.

Kemper, K.E., C.M. Reich, P.J. Bowman, C.J. Vander Jagt, A.J. Chamberlain, B.A. Mason, B.J. Hayes, and M.E. Goddard. 2015. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. Genet. Sel. Evol. 47:29. doi:10.1186/s12711-014-0074-4.


Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022

Khansefid, M., J.E. Pryce, S. Bolormaa, S.P. Miller, Z. Wang, C. Li et al. 2014. Estimation of genomic breeding values for residual feed intake in a multibreed cattle population. J. Anim. Sci. 92(8):3270–3283. doi:10.2527/jas.2014-7375.

Koivula, M., I. Strandén, G. Su, and E.A. Mäntysaari. 2012. Different methods to calculate genomic predictions–comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). J. Dairy Sci. 95(7):4065–4073. doi:10.3168/jds.2011-4874.

Legarra, A., O. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65. doi:10.1016/j.livsci.2014.04.029.

Legarra, A., and V. Ducrocq. 2012. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. J. Dairy Sci. 95:4629–4645. doi:10.3168/jds.2011-4982.

Lillehammer, M., T.H. Meuwissen, and A.K. Sonesson. 2011. Genomic selection for maternal traits in pigs. J. Anim. Sci. 89(12):3908–3916. doi:10.2527/jas.2011-4044.

Liu, Z., M.E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effects. J. Dairy Sci. 97(9):5833–5850. doi:10.3168/jds.2014-7924.

Lourenco, D.A.L., S. Tsuruta, B.O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J.K. Bertrand, T.S. Amen, L. Wang, D.W. Moser, and I. Misztal. 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. J. Anim. Sci. 93(6):2653–2662. doi:10.2527/jas.2014-8836.

Lund, M.S., A.P. Roos, A.G. Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 12:43. doi:10.1186/1297-9686-43-43.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter et al. 2009. Finding the missing heritability of complex diseases. Nature 461(7265):747–753. doi:10.1038/nature08494.

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

Meuwissen, T., and M. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185(2):623–631. doi:10.1534/genetics.110.116590.

Meuwissen, T.H.E., T. Luan, and J.A. Woolliams. 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J. Anim. Breed. Genet. 128(6):429–439. doi:10.1111/j.1439-0388.2011.00966.x.

Moen, T., J. Torgersen, N. Santi, W.S. Davidson, M. Baranski, J. Odegard et al. 2015. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in atlantic salmon. Genetics 200(4):1313–1326. doi:10.1534/genetics.115.175406.

Pryce, J., and B. Hayes. 2012. A review of how dairy farmers can use and profit from genomic technologies. Anim. Prod. Sci. 52:180–184. doi:10.1071/AN11172.

Sargolzaei, M., J.P. Chesnais, and F.S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478.

Schenkel, F. 2012. Inbreeding using genomics and how it can help. World Holstein Federation Conference, Toronto. http://www.whff.info/documentation/documents/2012_03_FlavioSchenkel.pdf.

van Binsbergen, R., M.C. Bink, M.P. Calus, F.A. van Eeuwijk, B.J. Hayes, I. Hulsegge, and R.F. Veerkamp. 2014. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. Genet. Sel. Evol. 46:41. doi: 10.1186/1297-9686-46-41.

Van Eenennaam, A.L., K.A. Weigel, A.E. Young, M.A. Cleveland, and J.C. Dekkers. 2014. Applied animal genomics: Results from the field. Annu. Rev. Anim. Biosci. 2:105–139. doi:10.1146/annurev-animal-022513-114119.

VanRaden. 2008. P. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

Vitezica, Z.G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. 93(5):357–366. doi:10.1017/S001667231100022X.

Weigel, K.A., P.C. Hoffman, W. Herring, and T.J. Lawlor, Jr. 2012. Potential gains in lifetime net merit from genomic testing of cows, heifers, and calves on commercial dairy farms. J. Dairy Sci. 95(4):2215–2225. doi:10.3168/jds.2011-4877.

Wiggans, G.R., P.M. Vanraden, and T.A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94(6):3202–3211. doi:10.3168/jds.2010-3866.

Wolc, A., H.H. Zhao, J. Arango, P. Settar, J.E. Fulton, N.P. O'Sullivan, R. Preisinger, C. Stricker, D. Habier, R.L. Fernando, D.J. Garrick, S.J. Lamont, and J.C. Dekkers. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. Genet. Sel. Evol. 47:59.

Yang, J., B. Beben, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.F. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large part of the heritability for human height. Nat. Genet. 42:565–569. doi:10.1038/ng.608.

## About the Authors

**Theo Meuwissen** is Professor in Animal Breeding and Genetics at the Norwegian University of Life Sciences, Ås, Norway. His research is directed at strategies for the use of new technologies in genetic improvement schemes. In the area of the design of breeding schemes, he developed optimal contribution selection: a selection method that maximizes the genetic improvement and at the same time controls the inbreeding in the population. The most fascinating new technology is the recent abundance of genomics data, and Meuwissen developed methods for the fine-scale mapping of genes combining linkage and linkage disequilibrium information. Together with Mike Goddard and Ben Hayes, he was the first to propose genomic selection: A selection method that effectively uses dense-genomic marker data for the prediction of genetic values of animals and enables the widespread use of genomics data in animal breeding. In the literature, this approach has been termed "the most promising application of molecular genetics in livestock populations." **Correspondence:** theo.meuwissen@nmbu.no.

**Dr. Ben Hayes** has a wide-ranging career designing genetic improvement programs for beef cattle, sheep, dairy cattle, and Atlantic salmon. His focus for the past 10 years has been the development of genomic selection methods. Recent work addresses the challenges of optimal breeding program design with genomic selection, incorporating whole-genome sequence information into genomic selection in the 1,000 bull genomes project, as well as improving feed conversion efficiency and heat tolerance of dairy cattle.

**Professor Michael Goddard** holds a joint appointment with the University of Melbourne and the Victorian Department of Economic Development, Jobs, Transport, and Resources as Professorial Fellow in Animal Genetics. His research interests are in quantitative genetics and their application to genetic improvement of livestock. Along with Theo Meuwissen and Ben Hayes, he invented a method, which has been called genomic selection or genomic prediction, to use genome-wide genetic markers to estimate the breeding value of individuals. He also collaborated with Jian Yang and Peter Visscher to resolve the so-called missing heritability paradox by showing that most polymorphisms affecting quantitative traits have such small effects that they do not reach statistical significance.

Downloaded from https://academic.oup.com/af/article/6/1/6/4638797 by guest on 22 January 2022