CSIRO PUBLISHING

Animal Production Science, 2012, **52**, 73–77
http://dx.doi.org/10.1071/AN11180

Review

Uses of genomics in livestock agriculture

M. E. Goddard

Department of Agriculture and Food System, University of Melbourne, Parkville, Vic. 3010, Australia and Biosciences Research Division, Department of Primary Industries, Bundoora, Vic. 3083, Australia and Cooperative Research Centre for Beef Genetic Technologies, Armidale, NSW 2351, Australia. Email: mike.goddard@dpi.vic.gov.au

Abstract. World demand for livestock products is likely to increase in coming decades but the cost of production could escalate faster than the price due to competition for land, water, grain and fertiliser and the effects of climate change and its mitigation. To remain competitive for these resources, livestock agriculture has to dramatically increase in efficiency of production. Genetic gain is one mechanism to achieve increased efficiency and there is the opportunity to utilise the scientific advances in genomics. Three ways in which genomics can be used are in additive genetic improvement, exploitation of non-additive genetic variance and management which exploits genotype by environment interactions to optimise management. Genomic selection is already being widely implemented in dairy cattle and beef cattle and sheep will follow in the future once the accuracy of genomic selection is high enough. The accuracy of equations that predict breeding value from DNA genotypes can be increased by increasing the size of the reference population from which the equations are estimated, increasing the density of markers, using genome sequences instead of markers, using more appropriate statistical procedures and incorporating biological information into the prediction. In the long term, genomic selection combined with reproductive technology that reduces the minimum age at breeding will greatly increase the rate of genetic gain. This will allow long-term increases in biological efficiency and short-term tailoring of livestock to meet the demands of particular markets and opportunities.

Received 17 August 2011, accepted 24 January 2012, published online 6 March 2012

Opportunity and challenges

Livestock agriculture faces great opportunities and challenges in the coming decades. As the world population increases to 9 billion by 2050 (FAO Commission on Genetic Resources for Food and Agriculture 2007) and as the incomes of many in developing countries increase, the demand for meat and dairy products will also increase (Delgado *et al*. 1999). However, there will also be increased competition for land, water, fertiliser and grain which will increase the cost of livestock production. Climate change will exacerbate these cost increases in three ways. First, it will increase the shortages of water, fertiliser and grain in parts of the world. Second, climate change will directly affect livestock in some places due to greater heat stress and tropical diseases (Hughes 2003). Third, it may lead to a charge for methane emission.

Livestock production could compete for scarce resources if the price of livestock products increased in line with input costs but this may not happen because consumers have alternatives to meat and dairy products and these will improve with time as food manufacturing technology improves.

Therefore livestock agriculture needs to increase in efficiency over the coming decades. There have been increases in productivity in the past, but these have been much faster in intensive industries such as poultry than in extensive ruminant production where they have just managed to maintain livestock agriculture as a competitive investment. The need for these gains will continue but will be increased by the competition for inputs discussed above, so we need much faster productivity gain in the future than in the past.

Genetic change in livestock is one source of gain in efficiency. The genomic revolution sweeping genetics and biology is providing new knowledge and new tools that can be applied to livestock production. For instance, the cost of genome sequencing has dropped to 1 millionth.

In the present paper, I will consider how genomics can be used to achieve faster genetic improvement in the dairy, beef and sheep industries. I will consider three types of genetic change – additive genetic improvement, exploitation of non-additive genetic variation and genotype by environment interaction (G × E).

Additive genetic improvement

Most traits of economic importance are quantitative or complex traits controlled by many genes and by environmental effects. Traditional improvement of these traits has relied on using phenotypic data and pedigrees to estimate the combined effect of all genes on the additive genetic value or breeding value of each animal. However, genetic variation is due to variation in the DNA sequence, so it would seem logical to select animals

carrying the most favourable alleles at all sites affecting profitability. This has been implemented at a small number of sites where mutations cause a large effect, such as those causing genetic abnormalities. Unfortunately, most of the polymorphic sites in the DNA that control economic traits are unknown. An alternative to selection on the causal sites is to use genetic markers linked to these causal sites. Such marker-assisted selection was relatively unsuccessful until the development of assays that could genotype thousands of single-nucleotide polymorphisms (SNPs) covering the entire genome. This made it possible to implement 'genomic selection' (Meuwissen *et al*. 2001) which is the use of a large panel of dense, genome-wide markers to predict the breeding value. Genomic selection relies on linkage disequilibrium (LD) between markers and causal polymorphisms which cause associations between markers and the traits that these causal polymorphisms affect. It is a statistical method that does not require identification of the genes or sites causing variation in the trait. It has been most successful in dairy cattle where the accuracy of predicting the breeding value from SNP genotypes is 0.7 in some cases (VanRaden *et al*. 2009).

The advantage of marker-assisted selection, including genomic selection, over traditional methods is greatest where traditional methods are difficult to implement (Meuwissen and Goddard 1996). This usually occurs because the phenotype of interest cannot be observed on selection candidates at the age when they can first be used for mating. For instance, milk production cannot be observed in bulls, meat tenderness cannot be measured on the live animal and adult wool production cannot be observed in yearling sheep. This limitation of traditional selection is more widespread than sometimes acknowledged. For instance, feed conversion efficiency is a major economic objective but it is seldom measured. Also, commercial production may take place in crossbred animals or in an environment different from that in which stud animals are selected.

The accuracy of genomic selection is not currently as high as that of a progeny test in most cases. The steps needed to raise the accuracy are described in Goddard *et al*. (2010) and summarised below.

The size of the reference population, that is the number of animals with phenotypes and genotypes from which the prediction equation is estimated, is critical (Goddard 2009). VanRaden *et al*. (2009) showed how accuracy increases with the size of the reference population. The effect of the reference population size interacts with the accuracy of the phenotype as a predictor of the breeding value. If the phenotype is a single measurement of the trait on the animal, this accuracy is the square root of the heritability. If the 'phenotype' is the average phenotype of a bull's daughters, then the accuracy is the correlation between the daughter average and the true breeding value of the bull. The accuracy of the prediction equations depends on Th^2, where T = number of animals in the reference population and h = accuracy of phenotype. At low levels of accuracy, accuracy squared is almost proportional to Th^2, but as accuracy increases, it reaches a plateau at 1.0 or less. Therefore, a bigger reference population is needed for individual measurements than for progeny test 'phenotypes' and a bigger population still is needed if the individual measurements are for a lowly heritable trait.

The value of Th^2 that is needed depends on the amount of LD in the population. Using the r^2 measure of LD, let M_e = 1/(average value of r^2 over all pairs of markers). Then the accuracy of genomic selection depends on Th^2/M_e (Goddard 2009). Thus, if LD is widespread, many pairs of markers show high values of r^2 and so M_e is low and Th^2/M_e is high, leading to high accuracy of genomic selection. M_e depends on the effective population size (N_e) because a small N_e leads to high LD.

In most breeds of livestock, recent N_e is small (100–200) and this causes widespread LD and helps genomic selection to work. By contrast, recent N_e in human populations is very large and it is difficult to predict even highly heritable traits such as height, with any accuracy from SNP genotypes (Lango Allen *et al*. 2010). However, if the population consists of a mixture of breeds, LD may be reduced. If the phase of LD varies among breeds, this means that the correlation (r) between a pair of markers may be positive in some breeds and negative in others, and so the average r over the breeds will be small and, therefore, r^2 will be small. Mixing breeds also generates some LD due to differences in allele frequency among breeds. But this source of LD is used to some extent in traditional genetic evaluation by fitting breed in the statistical model and so cannot be counted as an advantage of genomic selection unless the breed composition of animals in the population is unknown. LD phase varies among breeds of *Bos taurus* cattle unless the markers are very close together, say less than 10 kb apart (de Roos *et al*. 2008, 2009). Therefore, average r^2 declines towards zero as more breeds are added if the markers are 50 kb apart, but declines to a low but non-zero value if markers are 5 kb apart. This implies a large value of M_e, and hence a high value of Th^2 is needed for highly accurate genomic selection in a mixed-breed population.

The accuracy of genomic selection is higher in dairy cattle than in beef cattle and sheep (see papers Hayes *et al*. (2010) in these proceedings). This is not surprising, because in dairy cattle the size of the reference population (T) is larger than in beef and sheep, the phenotype is a daughter average and so h^2 is higher than in beef where individual phenotypes are frequently used, LD is high in some dairy breeds with small effective population size and prediction equations are based on data from one breed (e.g. Holstein) rather than a mixture of breeds, as is common in beef and sheep.

The accuracy of genomic selection also depends on the density of markers. The markers must be dense enough so that all QTL are in high LD with one or more markers. If this is not the case, some of the genetic variance due to QTL will not be detected by the markers. If the QTL have properties similar to the markers, then the proportion of genetic variance explained by the markers is $M/(M + M_e)$, where M is the number of markers (Goddard *et al*. 2011). If the QTL are unlike the markers, then the markers may not detect all the genetic variance, no matter how many markers are used. For instance, if the QTL have one rare allele they will not be in complete LD with any marker that has no equally rare allele. In humans, markers trace only 60% of the genetic variance for height (Yang *et al*. 2010). Use of haplotype information might overcome this problem, but has other disadvantages such as increasing the number of effects that must be estimated.

An alternative to genetic markers is to use the full genome sequence of each animal for the prediction of the breeding value. This is becoming possible due to the dramatic drop in the cost of genome sequencing and the ability to impute sequence from marker genotypes once a large reference panel of animals with full sequence exists for each species. Full sequence data provide the equivalent of very dense markers, and more importantly, they should include the causal polymorphisms. Meuwissen and Goddard (2010) showed that this led to a higher accuracy in predicting the breeding value than did marker panels.

The statistical method used to estimate breeding values from marker data affects the accuracy of the prediction. The best method to use depends on the genetic architecture of the trait (Hayes *et al*. 2010). For instance, if there are a huge number of QTL, each with a very small effect, then the method called BLUP is the best method. However, if only some markers are needed or if some have a larger effect, Bayesian methods perform better than BLUP. To get an advantage from genome-sequence data, it is probably necessary to use one of the Bayesian methods because, in sequence data, very few of the polymorphisms are expected to have an effect on the phenotype (Meuwissen and Goddard 2010).

Some of the Bayesian statistical methods, referred to in the previous paragraph, assume that only some of the markers or polymorphic sites in the sequence have an effect on the trait. That is, they can be described as a model selection process in which the analysis is identifying which markers have an effect and which markers have no effect. In full sequence data, the sites with an effect should be causal polymorphisms or other sites in complete LD with the causal site. The causal sites have a biological effect on the trait. They should be in a gene whose product affects the trait and they should be a site within the gene that affects either the structure of the gene product or its regulation. Among the 180 markers known to be associated with human height, some are close to genes with a known role in skeletal growth and some are in LD with a site that is known to affect the expression of the gene (Lango Allen *et al*. 2010). Therefore, the discoveries of associations between markers and traits can lead to new biological knowledge but we should also be able to use biological knowledge to improve the selection of sites to include in the equation to predict the breeding value. This approach should increase in power when we use genome sequence data combined with Bayesian statistical methods.

Simulation studies predict that the response to genomic selection will rapidly decline unless prediction equations are updated each generation (Muir 2007). This occurs because selection fixes the marker but not the QTL due to incomplete LD and because recombination erodes the LD (Goddard 2009). The problem is exacerbated if the prediction equation does not track QTL with rare favourable alleles that contribute disproportionately to the long-term selection response (Goddard 2009). However, these results may be too pessimistic because the very large number of QTL affecting most traits means that the changes in allele frequency will be slow and selection response will decline more slowly than predicted.

To maximise the long-term selection response, it will be important that genomic selection increases the frequency of rare favourable alleles. Traditional selection on phenotype does this automatically if inefficiently. Therefore, it may be desirable to include some selection on phenotype into a genomic selection program. This happens automatically when phenotypes are recorded on selection candidates and used in the calculation of the estimated breeding value (EBV). In addition, it would help if the prediction equation was estimated using phenotypes from selection candidates; that is, to have the stud animals form part of the reference population.

To maximise the benefit from genomic selection, it is often necessary to change the design of the breeding program. For instance, in dairy cattle one must drastically reduce generation length by selecting yearling bulls and heifers as sires and dams of the next generation. In traditional selection, phenotypes have to be recorded on selection candidates or their near relatives. This is not the case with genomic selection. It is possible to collect phenotypes and genotypes on commercial animals of unknown parentage, estimate prediction equations from this data and use them to select among a group of stud animals with no phenotypes. This design is advantageous because it allows phenotypes such as tenderness to be collected that are difficult to collect on stud animals. However, it has not been shown that this approach will work in practice and it conflicts with the suggestion made in the previous paragraph that phenotypes should be collected on stud animals. More research is required to find the best designs of breeding programs for genomic selection. However, it seems likely that the best designs will be very different from the industry structures of today.

Non-additive genetic variation

The value of commercial animals is determined by their phenotype, which depends on environmental effects and total genetic value. The difference between genetic value and breeding value is due to non-additive effects such as dominance and epistasis. The most common uses of non-additive effects are by exploiting heterosis and by the minimisation of inbreeding.

Inbreeding can be minimised in the short term by avoiding mating relatives based on knowledge of their pedigree. Homozygosity can be further reduced by knowledge of the marker genotypes; mates are chosen so as to minimise the homozygosity of the offspring (Pryce *et al*. 2012). This requires that both sires and dams have been genotyped and so genotyping needs to be inexpensive for this to be a profitable strategy in commercial sheep and cattle.

One aspect of inbreeding depression is increased incidence of homozygotes for recessive genetic abnormalities and lethals. The incidence of homozygous-affected offspring can be decreased in the same way as other inbreeding depression by avoiding mating sires and dams that are both carriers of the same abnormality. Note that culling carriers also decreases the incidence of affected young but this is an example of selection on breeding value and has the disadvantage that it competes with other criteria for culling.

Heterosis is greater, the more distantly related the breeds to be crossed are. Often the amount of heterosis in different crosses is already known from experiments, but if this is not the case, it could be predicted by estimating the relatedness of breeds on the basis of genetic markers (Goddard and Ahmed

1982). That is, breeds that are least closely related show the most heterosis when crossed.

If a crossbreeding program is already in place, the aim of selection within the parent lines should be to improve the crossbred offspring. Traditionally, this is carried out by reciprocal recurrent selection but this requires progeny testing and hence increases generation length. Markers could be used to predict the value of a parent's crossbred offspring by using a crossbred reference population. This should cause the heterosis in the cross to steadily increase.

If the value of different epistatic gene combinations were known (Carlborg and Haley 2004), this information could be used in mate allocation in the same way as described above for avoidance of inbreeding and recessive abnormalities.

Some genes show imprinting in which only the paternal or the maternal allele is expressed. This could be exploited by having separate sire and dam lines. For a gene that is only expressed when inherited from the sire, one would select for the desirable allele in the sire line but ignore this gene in the dam line.

In meat sheep, a structured, terminal crossbreeding program is common, but this is not the case in cattle. Given the new opportunity to select for crossbred performance, perhaps it is time to reconsider crossbreeding in dairy and beef. Alternatively, by forming a synthetic or composite breed, it might be possible to use genomic selection especially successfully to increase additive genetic merit.

Genotype by environment interaction (G × E)

If the breeding program is producing animals for a single environment or market, then the objective should be to improve breeding value for that objective. The G × E is important only if information is coming from outside the target environment. For instance, milk production in USA is a different, although correlated, trait to milk production in Australia. Consequently, phenotypic information from USA predicts the breeding value in Australia less accurately than the same phenotypic data from Australia. This inaccuracy caused by the G × E can be overcome by using genomic selection. For instance, a bull in the USA can have DNA collected and genotyped and used in the equation to predict the breeding value in Australia. Its EBV will then be as accurate as that of a bull in Australia with only DNA data. (This assumes that the LD between SNPs and causal polymorphisms in American Holsteins is the same as in Australian Holsteins, which is likely to be the case.) This should increase the internationalisation of breeding programs because an animal does not need close relatives, such as daughters, in every environment in which it might be selected for breeding.

Breeding programs often produce animals for more than one environment or market. For instance, cattle from the same breeding program might be sold for a domestic market and for an export market with different requirements. In this case, there are two alternative strategies. First, the breeding program can be split into parts that concentrate on only some of the environments or markets. Alternatively, animals bred by the single breeding program can be allocated to the environment or market in which they will be most profitable. For instance, a Charolais cattle breeder might sell bulls to beef producers in both southern and northern Australia and allocate bulls according to where they will be most profitable. Genomic selection is useful for this purpose because it is possible to evaluate the breeding value of a bull for performance in an environment other than that in which he has been raised.

This process of allocating animals to the most profitable environment or market could also be applied to commercial animals such as feedlot steers. For instance, the optimum number of days on feed could be decided for each individual steer. In this case, it is the future phenotypic value of the animal that is relevant, not its breeding value. The most accurate prediction of the future phenotype might utilise prediction of the breeding value, non-additive genetic value and environmental effects, for instance, by utilising the current phenotype as a predictor. Although there are benefits from allocating commercial animals to their most profitable environment, the benefit is limited to one animal and is not multiplied over many descendants and, consequently, the cost of deciding on the best environment must be small. However, provided the cost can be reduced, we could see a form of precision agriculture or personalised medicine applied to livestock in which each animal receives optimised treatment.

Breeding program design

When only phenotypic information is used to estimate the breeding value of an animal, the accuracy of this EBV increases as the animal ages. At first, the animal may have no phenotypic information of its own, but gradually the number of traits that can be measured increases and, eventually, this is supplemented by phenotypic information on progeny. Consequently, there is a trade-off in the design of the breeding program between accuracy of selection and generation length. However, DNA can be obtained at birth or even before and used in genomic selection to predict the breeding value. This prediction will not change as the animal ages, unless the prediction equation changes. Therefore, the optimum age for selection will usually be lower with genomic selection than with traditional phenotypic selection. For instance, dairy-cattle breeding programs are changing from using progeny-tested bulls to using yearling bulls.

Although it may be possible to make selection decisions at birth, it is not possible to obtain offspring from animals until a later age. Thus, technology that reduces the minimum age for breeding acts synergistically with genomic selection to increase the rate of genetic gain by decreasing the generation interval.

Under genomic selection, there is still a strong need for a database of animals with phenotypic measurements and genotypes from which to estimate prediction equations. However, there is less need than previously to measure phenotypes on selection candidates. This fact, combined with the additional value now possible from reproductive technology, could potentially change the design of breeding programs and even the industry structure that currently supports genetic improvement. It is not possible to predict what changes will occur, but over the next 20 years, I expect that there will be big changes in the genetic-improvement industry.

Conclusions

There is a strong need to increase the biological efficiency of livestock production to meet the rising costs of inputs expected in the future. Long-term genetic improvement, using genomics and reproductive technology, could achieve part of the increased efficiency needed. The increasing demand for milk and meat will come from non-traditional markets in developing countries and it is difficult to anticipate the exact requirements of this trade, other than low cost. Therefore, we will need methods to respond relatively quickly to take advantage of markets with specific requirements. Shorter-term genetic improvement, together with breed selection, mating plans and individual management, could generate cattle and sheep suited to these market opportunities.

References

Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nature Reviews. Genetics* **5**, 618–625. doi:10.1038/nrg1407

Delgado C, Rosegrant M, Steinfeld H, Ehui S, Courbois C (1999) Livestock to 2020 – the next food revolution. Food, agriculture and the environment, draft discussion paper 28. International Food Policy Research Institute, Food and Agriculture Organization of the United Nations and International Livestock Research Institute.

de Roos APW, Hayes BJ, Spelman RJ, Goddard ME (2008) Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503–1512. doi:10.1534/genetics.107.084301

de Roos APW, Hayes BJ, Goddard ME (2009) Reliability of genomic breeding values across multiple populations. *Genetics* **183**, 1545–1553. doi:10.1534/genetics.109.104935

FAO Commission on Genetic Resources for Food and Agriculture (2007) In 'The state of the world's animal genetic resources for food and agriculture'. (Eds B Rischkowsky, D Pilling) pp. 141–142. (FAO: Rome)

Goddard ME (2008) The use of high density genotyping in animal health. In 'Animal genomics for animal health'. (Eds M-H Pinard, C Gay, P-P Pastoret, B Dodet) pp. 383–389. (Developmental Biology Karger 132: Basel, Switzerland)

Goddard ME (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257. doi:10.1007/s10709-008-9308-0

Goddard ME, Ahmed AM (1982) The use of the genetic distance between cattle breeds to predict the heterosis in crosses. In 'Proceedings of the 2nd world congress on genetics applied to livestock production'. **8**, 377–382.

Goddard ME, Meuwissen THE, Hayes BJ (2010) Genomic selection in farm animal species – lessons learnt and future perspectives. In 'Proceedings of the 9th world congress on genetics applied to livestock production', Leipzig. Paper 0701.

Goddard ME, Hayes BJ, Meuwissen THE (2011) Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of Animal Breeding and Genetics* **128**, 409–421.

Hayes BJ, Pryce J, Chamberlain AJ, Goddard ME (2010) Genetic architecture of complex traits and the accuracy of genomic prediction: proportion of black in Holstein Friesian cattle as a model trait. *PLOS Genetics* **6**, e1001139. doi:10.1371/journal.pgen.1001139

Hughes L (2003) Climate change and Australia: trends, projections and impacts. *Austral Ecology* **28**, 423–443. doi:10.1046/j.1442-9993.2003.01300.x

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, *et al*. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838. doi:10.1038/nature09410

Meuwissen THE, Goddard ME (1996) The use of marker haplotypes in animal breeding schemes. *Genetics, Selection, Evolution.* **28**, 161–176. doi:10.1186/1297-9686-28-2-161

Meuwissen THE, Goddard ME (2010) Accurate prediction of genetic values for complex traits by whole genome resequencing. *Genetics* **185**, 623–631. doi:10.1534/genetics.110.116590

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.

Muir WM (2007) Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics* **124**, 342–355. doi:10.1111/j.1439-0388.2007.00700.x

Pryce JE, Hayes BJ, Goddard ME (2012) Novel strategies to minimise progeny inbreeding while maximising genetic gain using genomic information. *Journal of Dairy Science* **95**, 377–388.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* **92**, 16–24. doi:10.3168/jds.2008-1514

Yang J, Beben B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PF, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Missing heritability of human height explained by genomic relationships. *Nature Genetics* **42**, 565–569. doi:10.1038/ng.608