Advances in Animal Biosciences (2013), **4:3**, pp 618–625 © The Animal Consortium 2013
doi:10.1017/S2040470013000186

advances in animal biosciences

Genomic selection: the option for new robustness traits?

M. P. L. Calus[1†], D. P. Berry[2], G. Banos[3], Y. de Haas[1] and R. F. Veerkamp[1]

[1]*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 8200 AB Lelystad, The Netherlands;* [2]*Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Co. Cork, Ireland;* [3]*Animal & Veterinary Sciences Group, SRUC, Roslin Institute Building, Easter Bush, Penicuik, EH25 9RG Scotland, UK*

Genomic selection is rapidly becoming the state-of-the-art genetic selection methodology in dairy cattle breeding schemes around the world. The objective of this paper was to explore possibilities to apply genomic selection for traits related to dairy cow robustness. Deterministic simulations indicate that replacing progeny testing with genomic selection may favour genetic response for production traits at the expense of robustness traits, owing to a disproportional change in accuracies obtained across trait groups. Nevertheless, several options are available to improve the accuracy of genomic selection for robustness traits. Moreover, genomic selection opens up the opportunity to begin selection for new traits using specialised reference populations of limited size where phenotyping of large populations of animals is currently prohibitive. Reference populations for such traits may be nucleus-type herds, research herds or pooled data from (international) research experiments or research herds. The RobustMilk project has set an example for the latter approach, by collating international data for progesterone-based traits, feed intake and energy balance-related traits. Reference population design, both in terms of relatedness of the animals and variability in phenotypic performance, is important to optimise the accuracy of genomic selection. Use of indicator traits, combined with multi-trait genomic prediction models, can further contribute to improved accuracy of genomic prediction for robustness traits. Experience to date indicates that for newly recorded robustness traits that are negatively correlated with the main breeding goal, cow reference populations of ⩾10 000 are required when genotyping is based on medium- or high-density single-nucleotide polymorphism arrays. Further genotyping advances (e.g. sequencing) combined with post-genomics technologies will enhance the opportunities for (genomic) selection to improve cow robustness.

Keywords: genomic selection, robustness, dairy cattle

Genomic selection

Genomic selection is rapidly becoming the state-of-the-art selection methodology in dairy cattle breeding schemes around the world (Hayes *et al.*, 2009; Calus, 2010). The main benefit of genomic selection is that it can accurately identify genetically elite animals at a very young age, because it relies on estimates of genetic merit derived from DNA information (i.e. genomic breeding values). These genomic breeding values are considerably more accurate than the traditionally used pedigree indexes for young animals derived solely from information on the genetic merit of the animal's sire and dam. The increased accuracy of selection at a young age makes genomic selection attractive for dairy cattle breeding programs, where traditionally rapid annual genetic progress has been hindered by the long generation interval and low reproductive rate of cattle.

Because genomic selection is based on the evaluation of the DNA information of an individual, accurate knowledge of the effect of each variation in DNA on the range of performance traits evaluated in breeding programmes is required. The smallest variation at the DNA level is a single-nucleotide base change and these single-nucleotide polymorphisms (SNPs) are currently used in the implementation of genomic predictions within genomic selection procedures. Nowadays, dairy cattle genomic selection programmes use information on up to 54 001 SNPs on individual animals (Hayes *et al.*, 2009; Calus, 2010), although technology platforms with up to 777 962 SNPs are also commercially available (http://www.illumina.com).

The associations between each genotyped SNP and performance are estimated using a so-called reference population and confirmed in a different validation population (Figure 1). Accurate phenotypic and pedigree records are required in both populations. This reference population consists of several thousand animals with known genotypes and phenotypes. The initial concept of genomic selection was that the genotyped SNPs would be in linkage disequilibrium with the causal mutations (Meuwissen *et al.*, 2001), or in other words a given allele of a genotyped SNP

† E-mail: mario.calus@wur.nl

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press

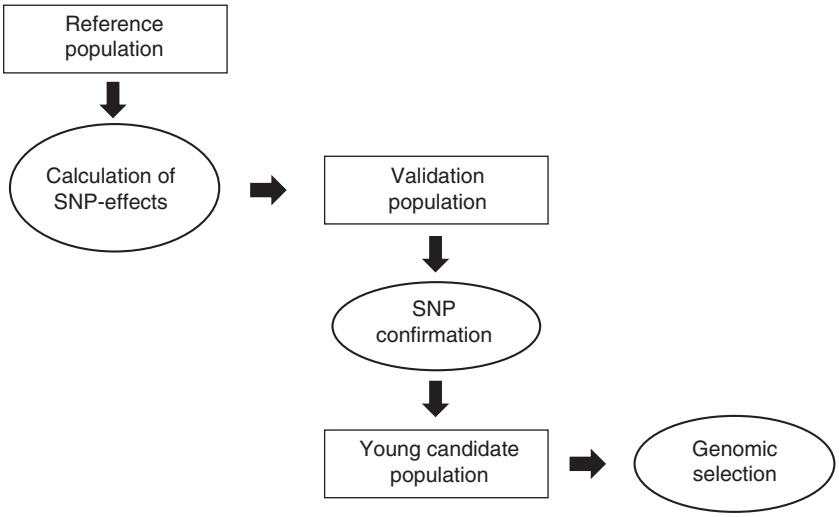


Figure 1 Schematic illustration of the populations and processes of genomic selection.

was more often than not co-inherited with a given allele in the causal mutation, affecting the traits in the breeding goal/selection index. However, several studies have shown that (the accuracy of) genomic selection relies heavily on the relationships between the reference population and the selection candidates (Habier *et al.*, 2007; Pszczola *et al.*, 2012a); stronger genetic relationships are associated with increased accuracies. An optimum balance of animal relationships within the reference population *v.* relationships between the reference, validation and selection candidate populations is also crucial. At the same time, care should be taken in the validation process to avoid an overestimation of the expected accuracy. Such an overestimation may be owing to results from the same genetic evaluation being used for both reference and validation animals (Amer and Banos, 2010). The problem is exacerbated with low-heritability traits.

In addition, genomic selection accuracy is a function of the size of the reference population, the heritability of the trait analysed and the effective size of the overall population from which the reference population is sampled (Daetwyler *et al.*, 2008). This has implications for genomic selection for robustness traits as it implies that a large genotyped population (i.e. at least several thousand animals) is required, which includes animals that are also phenotyped for animal robustness. All else being equal, the lower the heritability of the trait under selection, the greater the reference population required. In this paper, we describe the potential of using genomic selection for traits related to cow robustness, some of which may be lowly heritable and difficult to record, and the prospects and challenges involved in this application.

Animal robustness

A commonly used definition of robustness of a cow is 'the ability to maintain homeostasis in commonly accepted and sustainable dairy herds in the near future' (Ten Napel *et al.*, 2009). The term robustness is generally used in relation to non-production or so-called functional traits. In the last few decades, dairy cattle breeding goals have broadened from primarily focusing on production traits to now including both production and functional traits in a balanced breeding goal (Miglior *et al.*, 2005; Shook, 2006). Figure 2 gives a schematic representation of many (groups of) traits related to robustness that are currently included in international breeding goals, together with the actual traits that are measured for management purposes and genetic improvement. Several of these functional traits are measured on a large scale, such as fertility-related interval traits, somatic cell count and many conformation traits. Nonetheless, most of these traits are, in fact, either indicator traits or index traits (Figure 2) for other underlying traits, which are usually difficult or expensive to measure in a sufficiently large population to undertake accurate genetic evaluations. Examples are calving interval used as a measure of fertility, somatic cell count to predict (susceptibility to) mastitis, and feet and leg conformation traits to predict mobility and lameness. Each of these indicator traits themselves are complex phenotypes of underlying physiological phenotypes. For example, calving interval is made up of a combination of the ability of the cow to ovulate post calving, express oestrus, conceive and establish pregnancy, as well as gestation length. Measuring these underlying traits in a traditional dairy cattle breeding scheme is usually not an option, as the costs to obtain records on hundreds of daughters on commercial farms for each male selection candidate are simply too high.

New low-cost phenotyping strategies are currently under investigation to better predict or measure these difficult to acquire phenotypes associated with animal robustness traits. One of these strategies, investigated as part of the RobustMilk project, is the use of mid-IR spectroscopy of milk to predict energy intake and energy balance (McParland *et al.*, 2011). Image analysis techniques and sensor technology will also lead to the acquisition of larger number of low-cost phenotypes in the future (Berry *et al.*, 2013).

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press

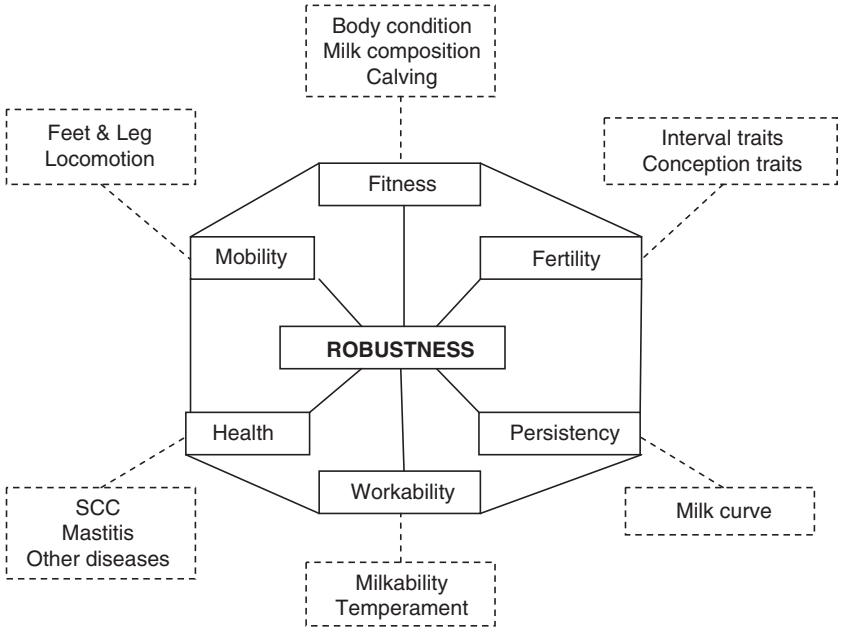


Figure 2 Dissecting robustness in dairy cattle; component functions in solid boxes; recorded traits associated with each function in dashed boxes.

Compiling reference populations for robustness traits

Genomic selection somewhat relaxes the requirement to have phenotypic measurements on many progeny for each selection candidate, which makes it especially attractive for difficult or expensive to measure traits. This is because genomic selection opens up the opportunity to establish specialised reference populations of limited size in which these novel or difficult to measure traits are recorded without the need of progeny of the selection candidates being present. These reference populations may be commercially owned (e.g. nucleus-type herds) to provide commercial advantage, nationally owned funded by public good, or jointly established international reference populations with strong genetic linkages to the selection candidates of the respective countries. Moreover, already collected data from (international) research experiments or research herds (Wall *et al.*, 2011; Banos *et al.*, 2012) could also be pooled. Within the RobustMilk project, international data were collated for progesterone-based traits (Berry *et al.*, 2012), feed intake and energy balance-related traits (Veerkamp *et al.*, 2012b), and somatic cell score (Wijga *et al.*, 2012). Similar pooling of data for feed intake and efficiency in dairy cattle has been undertaken between Australia and New Zealand (Pryce *et al.*, 2012), and between Australia and Europe (de Haas *et al.*, 2012). Other initiatives are underway to pool data for feed intake (Veerkamp *et al.*, 2012a) and methane emission (de Haas *et al.*, 2011). There is potential for substantial increases in the power of detecting single effects (e.g. SNP) on such traits when individual compatible data sets are merged (Banos *et al.*, 2012).

The objective of the RobustMilk project was to combine data from individual partners recognising that each individual partner had too small a reference population to generate highly accurate genomic predictions. Within the RobustMilk project, de Haas *et al.* (2012) compared the accuracy of genomic prediction using only the individual country data sets, *v.* using an overall combined data set. In that study, data sets for dry matter intake were combined across countries (Australia, United Kingdom and The Netherlands), and the pairwise genetic correlations for dry matter intake between countries varied from 0.36 to 0.74. De Haas *et al.* (2012) showed that (1) a multi-trait model achieved the highest gain in accuracy across all countries and (2) all countries benefitted when all countries were included together in the prediction model. In a few situations when only two countries were included, however, realised accuracies in fact were lower compared with a scenario where only data from the country itself were used. These results indicate that generally there is an expected benefit of combining data, albeit in each specific case compatibility of data sets needs to be investigated. Compatibility can be established by empirical quantification of the realised benefit not only in terms of prediction accuracy, but can also be assessed at the level of the data by estimating genetic correlations between data sets and by investigating genetic links between data sets, either through pedigree or SNP information. There are no straightforward guidelines on how strong these links should be. Nevertheless, it is worth noting that problems of limited connectedness may be minimised using SNP information instead of pedigree information (Pszczola *et al.*, 2012a), because the SNP information provides a more accurate reflection of the relationships between pairs of animals, including pairwise relationships that are zero in a relationship matrix calculated from pedigree.

When a reference population needs to be established for difficult to measure traits, the most likely cost-effective strategy is to generate genotypes and phenotypes on cows,

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press

instead of genotyping bulls and phenotyping large paternal half-sib progeny groups (Buch *et al.*, 2012; Van Grevenhof *et al.*, 2012). On the basis of cow reference populations consisting of 600 to 2000 cows with single phenotypic measurements, the accuracy of direct genomic values (DGV) for traits with heritabilities ranging from 0.2 to 0.5 was reported to be in the range of 0.3 to 0.5 (Verbyla *et al.*, 2010; de Haas *et al.*, 2011 and 2012), which closely resembles the theoretical expected range of 0.29 to 0.63 based on a deterministic formula (Daetwyler *et al.*, 2008; Daetwyler, 2009).

An important question is which animals should be included in the optimal construction of a reference population. With regard to performance of the cows, it has been shown that sampling extreme cows (i.e. from both tails of the distribution of phenotypes) yields higher accuracy than selecting cows at random or only selecting the best cows (Jiménez-Montero *et al.*, 2012). Selecting the best cows, in fact, yielded the lowest accuracy. Many studies have shown that the degree of relationship between reference populations and selection candidates affects the prediction accuracy (Habier *et al.*, 2007; Pszczola *et al.*, 2012a). These relationships are especially important for small reference populations (Wientjes *et al.*, 2013). In addition, it has been shown that strong relationships among animals in the reference population, in fact, have a negative effect on the average accuracy of genomic predictions in selection candidates (Pszczola *et al.*, 2012a). As a consequence, the optimal reference population design maximises the relationships between the reference population and the evaluated animals, while minimising the relationship among animals in the reference population.

Both from the perspective of relationship to the evaluated animals and from the perspective of sampling extreme phenotypes, it is likely that adding 'unique' animals to the reference population leads to higher increases in accuracy compared with adding animals at random. This implies that for traits where phenotypes are abundant, animals could be selected on the basis of extreme phenotypes, that is, selecting the best and the worst animals. For traits where phenotypes are difficult or expensive to obtain, or in future scenarios where the majority of animals in a population are genotyped, entire populations could be screened on the basis of genotypes to select an optimal set of animals that needs to be phenotyped and subsequently included in the reference population, because they optimally contribute to the accuracy of genomic prediction in the whole population.

Implementing the aforementioned theoretical optimum design of a reference population may of course not be possible because of practical limitations. To achieve sufficiently high accuracy of genomic prediction, several alternative strategies exist. Phenotypic data may be available from past experiments, but no DNA of these animals may be available. However, those records can be used in the analysis, using either relationship matrices that combine genomic and pedigree-based relationship matrices (Aguilar *et al.*, 2010; Veerkamp *et al.*, 2011), or using sophisticated imputation algorithms to derive their genotypes (Hickey *et al.*, 2012).

Predicted accuracy of genomic selection for robustness traits

In recent years, several simulation studies have investigated the potential benefit of genomic selection for difficult or expensive to measure traits. Buch *et al.* (2012) quantified the added benefit of using phenotypic records of cows, in addition to sires and test bulls with progeny, to enable genomic predictions for this trait. They concluded that cow phenotypes are especially important when the reference population is small, and that DGV accuracies reach acceptable levels to begin selection within 3 years after starting recording the new trait. Amer (2012) investigated the effect on genetic gain by switching from a progeny-testing scheme to genomic selection, both on production and robustness traits that were previously selected based on progeny testing. He showed that genomic selection, in fact, may result in a selection response favouring production traits relative to robustness traits owing to an increased difference in the accuracy of selection achieved for robustness compared with production traits when implementing genomic selection. A similar trend was shown for fitness *v.* production traits by Egger-Danner *et al.* (2012). Those authors, however, also showed that starting to include direct health traits in the index did result in a similar increase of genetic gain across trait groups when introducing genomic selection. Thus, the problem of favouring production traits in genomic selection at the expense of, for example, robustness traits can be reduced by increased recording of the robustness traits to improve the accuracy of selection for those traits, or by improving the definition or statistical modelling of the trait, thereby potentially increasing the heritability of the trait.

Genomic breeding values of robustness traits may still have, at most, moderate accuracy depending on the nature of the trait because the heritability may be low and the reference populations used may be limited in size. We theoretically predicted the required reference population size to achieve genomic breeding values with a reliability of 30%, 50% or 70%. Details on the predictions are presented in the Appendix. The required reference populations to achieve the desired reliabilities are presented in Figure 3, assuming either a cow reference population with one record per cow (Figure 3a), or a bull reference population, where the phenotype of each bull is based on 100 daughters with one record each (Figure 3b). Figure 3b shows that a reference population of $>$10 000 bulls, which is currently available in several countries (Lund *et al.*, 2011), yields reliabilities of DGVs of at least 70% for traits with a heritability $>$0.1. Figure 3a shows that a reference population of 2000 cows yields reliabilities of 30% for a trait with a moderate heritability of 0.3. This implies that genomic selection based on a limited cow reference population will yield genomic breeding values with reliabilities that are much lower than currently accepted in the dairy cattle industry. Marketing individual bulls based on breeding values with such low reliabilities is unlikely to happen. Nevertheless, achieving genetic progress for traits at the level of the whole population,

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press

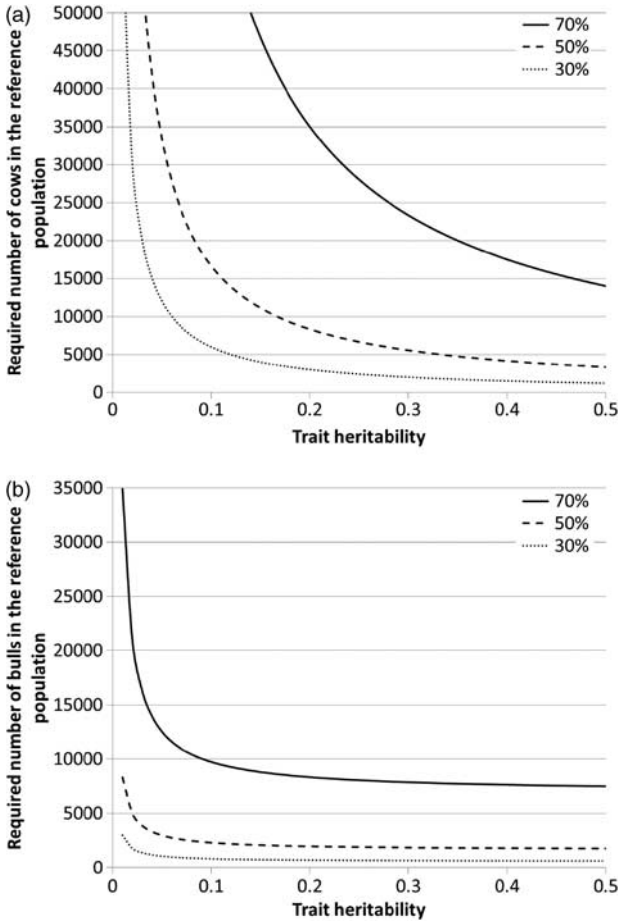


Figure 3 Required number of cows (a) or bulls (b) in the reference population to achieve a reliability of 30%, 50% or 70% as a function of trait heritability. *x*-axis = trait heritability; *y*-axis = number of animals.

on the basis of cow reference populations of 2000 to 10 000 cows, is possible (Calus *et al.*, 2013).

Multi-trait genomic prediction – combining old and new successful strategies

The strategy to use less-expensive predictor traits in multi-trait prediction models proved to be successful in traditional breeding programmes to increase the accuracy of prediction. This therefore suggests that the accuracy of genomic prediction may also benefit from using predictor traits in combination with multi-trait analysis. So far, only a few simulation studies have investigated this combined model, and those studies indeed support that multi-trait genomic prediction can lead to a considerable increase in genomic prediction accuracy (Calus and Veerkamp, 2011; Jia and Jannink, 2012).

Two studies have investigated the benefit of using multi-trait genomic prediction for US Holsteins. Using multi-trait genomic prediction for first-parity conception rate resulted in a doubling of the reliability (prediction accuracy squared) of genomic predictions (Aguilar *et al.*, 2011). Similarly, it was shown that the reliability of genomic breeding values for the conformation trait 'strength' increased from 0.40 to 0.45 when a multi-trait genomic prediction model with 18 conformation traits was used instead of a single-trait model (Tsuruta *et al.*, 2011). Both studies used the so-called 'single-step approach' that combines genotype and pedigree information of genotyped and ungenotyped animals in a single relationship matrix, and considered scenarios where genotyped animals generally had phenotypes for all traits included in the model. Another study, using data generated within the RobustMilk project, investigated the additional benefit of exploiting a multi-trait GBLUP-type model to predict genomic breeding values for dry matter intake, using measurements for milk yield and live weight as indicator traits (Pszczola *et al.*, 2012b). This study showed that indeed the indicator traits improved the accuracy of prediction for dry matter intake and also indicated that the accuracy was similar to a multi-trait pedigree-based model. This was most likely the result of moderate to strong genetic correlations between the predicted trait and the indicator traits, as it has been shown previously that added benefit of genomic information in multi-trait models is decreased when indicator traits with strong genetic correlations are used (Calus and Veerkamp, 2011).

Within the RobustMilk project, a Bayesian model was developed that can use data from two traits that are each measured on a separate group of animals (Calus *et al.*, 2012). This model was applied to a scenario where the one group of animals was a cow reference population and the other group of animals was a bull reference population. Results showed that accuracies of genomic prediction for the trait measured on the cows benefitted from exploiting the additional information on the bull trait and helped to reduce potential bias in predicted breeding values. In addition, using the cow and bull data combined resulted in increased power to detect quantitative trait loci (QTL), using the Bayesian model. Similar results were found when using this model in a genome-wide association study for milk fatty acid content as part of the RobustMilk project (Bastin *et al.*, 2012). In addition, evidence for two QTLs related to progesterone levels, detected using the single-trait model, improved when a bivariate model was applied using information on correlated fertility traits (Berry *et al.*, 2012).

Predicted accuracy and response to genomic selection, using a multi-trait approach with one trait recorded on cows and the other on bulls, was investigated using deterministic simulation within the RobustMilk project (Calus *et al.*, 2013). Scenarios were considered where phenotypes for a new trait were available on up to 10 000 cows, whereas a bull reference population of up to 20 000 animals was available for an overall index, which had a genetic correlation of −0.5, 0.0 or 0.5 to the new trait. When the new trait had a negative correlation with the overall index, achieving favourable genetic progress for the new trait was only possible in the extreme situation where the new trait had a moderate heritability (0.30) and when it was economically as important as the overall index. In all other scenarios with a negative correlation between the new trait and the overall index,

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press

the cow reference population was not sufficient to achieve favourable genetic progress for the new trait. When the new trait had an economic value that was at least 20% of the economic value for the overall index, using the cow reference population did diminish the negative genetic response for the new trait caused by selection for the overall index. In all scenarios where the genetic correlation between the new trait and the overall index was zero or positive, considerable genetic response could be achieved using a reference population of up to 10 000 cows, the rate of response achievable depending on the heritability and the economic value of the new trait.

Future perspectives for genomic selection for robustness traits

Genotyping all newborn (female) calves in dairy cattle populations as part of routine herd management will soon become the norm as the ever-decreasing cost of genotyping continues. In this scenario, genotype data will completely replace pedigree information in genetic evaluations. The major benefit of such a scenario is that animals to be phenotyped and thereafter included in the reference population can be strategically selected on the basis of their genotypes, to design a reference population that is optimally related to the whole population. In addition, wide-spread genotyping enables application of 'genomic mating plans', that is, mating plans that are based on complementarity of the genotypes between mated animals. Such genomic mating plans can be used to control or reduce (genomic) inbreeding levels. Minimising inbreeding is particularly important for robustness traits because inbreeding depression is greatest for traits associated with animal fitness (McParland *et al.*, 2007). Moreover, genomic mating plans can be potentially used to exploit non-additive effects (Toro and Varona, 2010), which is particularly important, given the proliferation of crossbreeding in some populations. An important issue for difficult or expensive to measure traits is the persistence of the accuracy of genomic predictions across generations. Persistency of accuracy across generations reduces the required number of additionally phenotyped animals that need to be added per generation to retain the prediction accuracy at the same level. One way to achieve this is to increase the degree of co-inheritance between the genotyped marker and the true causative mutation. This can be achieved through whole genome sequencing. These high-density data may be obtained on the reference population directly, or be imputed using information of other individuals in the same population. However, considerable attention must be given to developments in algorithms to optimally exploit the plethora of genotype information emanating from sequence data.

A phenotype observed in the field (e.g. fertility) is a direct consequence of the additive genetic makeup of the animal as well as non-additive genetic (e.g. interactions between genes within biological pathways) and other environmental influencing including errors (e.g. cow inseminated at the inappropriate time of the oestrus cycle). Being able to dissect complex phenotypes like robustness into its individual components may increase the heritability of these components (i.e. less residual noise) and aid in resolving antagonisms between traits (e.g. calving interval and pregnancy fertility traits), and thereby increase genetic gain. Omic technologies, such as (targeted) transcriptomics, metabolomics and proteomics, could provide useful tools in dissecting robustness, if the costs are sufficiently low and if the technologies can be applied to relatively large populations of animals.

Conclusions

Genomic selection opens up opportunities to begin genetic selection for traits where generating a large population of phenotyped animals is currently prohibitive. Strategic genotyping and phenotyping, especially through joint (international) partnerships can reduce the cost per partner of generating accurate genomic predictions for robustness traits. The EU project RobustMilk has set an example on how to establish a multi-country reference population for scarcely recorded traits, and greatly contributed to insights on how to start genomic selection for robustness traits. Experiences to date indicate that for newly recorded robustness traits that are negatively correlated with the main breeding goal, cow reference populations of $\geqslant$10 000 are required when genotyping is based on medium- or high-density SNP arrays. Further genotyping advances (e.g. sequencing) combined with post-genomic technologies will enhance the opportunities for selection to improve cow robustness.

Acknowledgements

The authors acknowledge funding from the RobustMilk project. The RobustMilk project is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708. RobustMilk acknowledges the financial support of EAAP towards the cost of publication. The content of this paper is the sole responsibility of the authors and it does not necessarily represent the views of the Commission or its services.

For more information, please visit the website www.robustmilk.eu, read the other papers in these proceedings or contact the project coordinator: Roel F. Veerkamp (E-mail: Roel.Veerkamp@wur.nl).

References

Aguilar I, Misztal I, Tsuruta S, Wiggans GR and Lawlor TJ 2011. Multiple trait genomic evaluation of conception rate in Holsteins. J Dairy Sci 94, 2621–2624.

Aguilar I, Misztal I, Johnson D, Legarra A, Tsuruta S and Lawlor T 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci 93, 743–752.

Amer PR 2012. Turning science on robust cattle into improved genetic selection decisions. Animal 6, 551–556.

Amer PR and Banos G 2010. Implications of avoiding overlap between training and testing data sets when evaluating genomic predictions of genetic merit. J Dairy Sci 93, 3320–3330.

Banos G, Coffey MP, Veerkamp RF, Berry DP and Wall E 2012. Merging and characterising phenotypic data on conventional and rare traits from dairy cattle experimental resources in three countries. Animal 6, 1040–1048.

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press

Bastin C, Gengler N, Soyeurt H, McParland S, Wall E and Calus MPL 2012. Genome-wide association study for milk fatty acid composition using cow versus bull data. In Proceedings of the 63rd Annual Meeting of the EAAP, Bratislava, Slovakia, p. 93.

Berry DP, McParland S, Bastin C, Wall E, Gengler N and Soyeurt H 2013. Phenotyping of robustness and milk quality. Advances in Animal Biosciences 4, 600–605.

Berry DP, Bastiaansen JWM, Veerkamp RF, Wijga S, Wall E, Berglund B and Calus MPL 2012. Genome-wide associations for fertility traits in Holstein–Friesian dairy cows using data from experimental research herds in four European countries. Animal 6, 1206–1215.

Buch LH, Kargo M, Berg P, Lassen J and Sorensen AC 2012. The value of cows in reference populations for genomic selection of new functional traits. Animal 6, 880–886.

Calus MPL 2010. Genomic breeding value prediction: methods and procedures. Animal 4, 157–164.

Calus MPL and Veerkamp RF 2011. Accuracy of multi-trait genomic selection using different methods. Genetics Selection Evolution 43, 26.

Calus MPL, De Haas Y and Veerkamp RF , 2012. Increasing accuracy of genomic prediction combining cow and bull reference populations. In Proceedings of the 4th International Conference on Quantitative Genetics, Book of Abstracts, Edinburgh, p. 136.

Calus MPL, de Haas Y, Pszczola M and Veerkamp RF 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. Animal 7, 183–191.

Daetwyler HD 2009. Genome-wide evaluation of populations. PhD thesis, Wageningen University, Wageningen.

Daetwyler HD, Villanueva B and Woolliams JA 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 3, e3395.

de Haas Y, Windig JJ, Calus MPL, Dijkstra J, de Haan M, Bannink A and Veerkamp RF 2011. Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. J Dairy Sci 94, 6122–6134.

de Haas Y, Calus MPL, Veerkamp RF, Wall E, Coffey MP, Daetwyler HD, Hayes BJ and Pryce JE 2012. Improved accuracy of genomic prediction for dry matter intake of dairy cattle from combined European and Australian data sets. J Dairy Sci 95, 6103–6112.

Egger-Danner C, Willam A, Fuerst C, Schwarzenbacher H and Fuerst-Waltl B 2012. Hot topic: effect of breeding strategies using genomic information on fitness and health. J Dairy Sci 95, 4600–4609.

Goddard M 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245–257.

Habier D, Fernando R and Dekkers J 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177, 2389–2397.

Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME 2009. Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92, 433–443.

Hickey J, Kinghorn B, Tier B, van der Werf J and Cleveland M 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. Genetics Selection Evolution 44, 9.

Jia Y and Jannink J-L 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192, 1513–1522.

Jiménez-Montero JA, Gonzalez-Recio O and Alenda R 2012. Genotyping strategies for genomic selection in dairy cattle. Animal 6, 1216–1224.

Lund M, de Roos S, de Vries A, Druet T, Ducrocq V, Fritz S, Guillaume F, Guldbrandtsen B, Liu Z, Reents R, Schrooten C, Seefried F and Su G 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genetics Selection Evolution 43, 43.

McParland S, Kearney JF, Rath M and Berry DP 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein–Friesians. J Dairy Sci 90, 4411–4419.

McParland S, Banos G, O'Donovan M, Coffey M, MCCarthy B, O'Neill B, Wall E and Berry D 2011. Genetic evaluations for energy balance a real possibility? In Proceedings of the Interbull Open Meeting, Interbull, Stavanger, Norway, pp. 255–259.

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

Miglior F, Muir BL and Van Doormaal BJ 2005. Selection indices in Holstein cattle of various countries. J Dairy Sci 88, 1255–1263.

Pryce JE, Arias J, Bowman PJ, Davis SR, Macdonald KA, Waghorn GC, Wales WJ, Williams YJ, Spelman RJ and Hayes BJ 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J Dairy Sci 95, 2108–2119.

Pszczola M, Strabel T, Mulder HA and Calus MPL 2012a. Reliability of direct genomic values for animals with different relationships within and to the reference population. J Dairy Sci 95, 389–400.

Pszczola M, Veerkamp RF, De Haas Y, Strabel T and Calus MPL , 2012b. Predictor traits improve accuracy of genomic breeding values for scarcely recorded traits. In Proceedings of the 4th International Conference on Quantitative Genetics, Book of Abstracts, Edinburgh, p. 151.

Shook GE 2006. Major advances in determining appropriate selection goals. J Dairy Sci 89, 1349–1361.

Ten Napel J, Calus MPL, Mulder HA and Veerkamp RF 2009. Genetic concepts to improve robustness of dairy cows. In Breeding for robustness in cattle (eds. M Klopčič, R Reents, J Philipsson and A Kuipers), EAAP publication no. 126, pp. 35–44. Wageningen Academic Publisher, Wageningen, The Netherlands.

Toro MA and Varona L 2010. A note on mate allocation for dominance handling in genomic selection. Genetics Selection Evolution 42.

Tsuruta S, Misztal I, Aguilar I and Lawlor TJ 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J Dairy Sci 94, 4198–4204.

Van Grevenhof E, Van Arendonk J and Bijma P 2012. Response to genomic selection: the Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. Genetics Selection Evolution 44, 26.

Veerkamp RF, Calus MPL and De Haas Y 2012a. Selection for feed intake in dairy cattle using genomic selection. In Proceedings of the ICAR 2012 Conference in Cork, Ireland, pp. 34–35.

Veerkamp RF, Mulder HA, Thompson R and Calus MPL 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. J Dairy Sci 94, 4189–4197.

Veerkamp RF, Coffey MP, Berry DP, de Haas Y, Strandberg E, Bovenhuis H, Calus MPL and Wall E 2012b. Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries. Animal 6, 1738–1749.

Verbyla KL, Calus MPL, Mulder HA, de Haas Y and Veerkamp RF 2010. Predicting energy balance for dairy cows using high-density single nucleotide polymorphism information. J Dairy Sci 93, 2757–2764.

Wall E, Coffey M, Veerkamp RF, McParland Sand and Banos G , 2011. Lessons learned in pooling data for reference populations. In Proceedings of the Interbull Open Meeting, Interbull, Stavanger, Norway, pp. 12–18.

Wientjes YCJ, Veerkamp RF and Calus MPL 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193, 621–631.

Wijga S, Bastiaansen JWM, Wall E, Strandberg E, de Haas Y, Giblin L and Bovenhuis H 2012. Genomic associations with somatic cell score in first-lactation Holstein cows. J Dairy Sci 95, 899–908.

Appendix

The reliability of a genomic breeding value (r^2) can be predicted using the general formula (Daetwyler *et al.*, 2008; Daetwyler, 2009):

$$r^2 = q^2 \frac{\frac{n_p}{n_G} h^2}{\frac{n_p}{n_G} h^2 + 1} \quad (1)$$

where q^2 is the proportion of the total genetic variance captured by the markers, n_p is the number of phenotypic records used (assuming one record per animal), n_G is the number of QTL or effective chromosome segments and h^2 is the heritability of the trait. For q^2, we used a value of 0.8, which is reported for the commonly used 50k SNP chip (Daetwyler, 2009). The chosen value for n_G was 1000 to reflect the effective number of independent chromosomal segments (Goddard, 2009).

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press

For bull reference populations, where the phenotype of individual bulls is their (deregressed) breeding value based on daughter information, h^2 in equation (1) can be replaced by the reliability of the bulls' daughter-based breeding value. This reliability can be calculated as

$$r^2 = \frac{\frac{1}{4}nh^2}{1 + \frac{1}{4}(n-1)h^2} \quad (2)$$

where n is the number of daughters (assumed to be 100). To calculate the number of records required to obtain a certain reliability r^2, equation (1) was rearranged to

$$n_p = \frac{r^2 n_G}{h^2(q^2 - r^2)} \quad (3)$$

https://doi.org/10.1017/S2040470013000186 Published online by Cambridge University Press