Editorial: Genomic selection with numerically small reference populations

M. P. L. Calus[†]

Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 338, 6700 AH Wageningen, The Netherlands

Genomic selection (GS) is becoming the state-of-the-art selection technology in many breeding programmes around the world. GS relies on direct genomic values (DGV), also often termed genomic estimated breeding values, that are estimated from a reference or training population with known genotypes and phenotypes. The process of estimating DGV is known as genomic prediction. Two main factors that determine the accuracy of GS are the size of the reference population (Meuwissen et al., 2001) and the relationship between the reference population and the predicted animals (Pszczola et al., 2012). It has been shown that there is an interaction between both factors; when the reference population is small it is especially important to achieve a high relationship between the reference population and the predicted animals (Wientjes et al., 2013).

Obtaining a sufficiently large reference population, typically comprising at least a few thousand animals, may be a challenge for numerically small populations and for breeding programmes with limited resources. Centred around the topic of improving the accuracy of GS for small populations, seven papers are presented here that build on presentations during the 10th World Conference of Genetics Applied to Livestock Production in August 2014 in Vancouver, Canada.

The first paper provides a review of theory and practical results on strategies and methods to increase the accuracies of DGV in numerically small dairy populations, by increasing the reference population by adding animals from other reference populations (Lund et al., 2016). When added animals originated from the same breed, considerable gains can be expected and the most promising strategy was reported to be increasing the relationship between genotypes of predicted animals and phenotypes in the reference populations. When added animals originated from a different breed, expected gains diminish when the added breed becomes more distantly related and the most promising strategy seems to be to gear models and strategies to focus genomic predictions on markers closer to the causative variants. The second paper extends on this by reviewing strategies to develop sufficiently accurate genomic predictions for hard to measure traits in dairy cattle that typically also are faced with limited number of phenotypes (Schöpke and Swalve, 2016). In particular, the potential benefit of contract herds that employ extensive phenotyping is emphasized.

In dairy cattle, one commonly envisaged way to increase the size of the reference population is to extend common bull reference populations with genotyped cows. Koivula et al. (2016) showed a 2% to 4% unit increase in reliability for production traits, when adding up to 7143 genotyped cows to a reference population comprising of 4413 bulls. As indicated by Koivula et al. (2016), it should be avoided to only add genotyped bull dams and their genotyped daughters, because this represents a selected group of animals and created some bias in their single-step evaluation.

International exchange of information has been common practice through Interbull for several decades. The relative importance of international information was often limited when evaluations relied on pedigree information, owing to limited connectedness between different (groups of) countries. With GS, genetic connectedness is much less of an issue, because in principle all animals within a breed can have some genomic relationship with each other. To capitalize on this, Su et al. (2016) investigated the potential to improve genomic predictions in Danish Jerseys relying on 1050 to 1250 Danish bulls, by adding 1150 US Jersey bulls, 4800 Danish Jersey cows or both. On average, adding the Danish Jersey cows was slightly more advantageous than adding the US Jersey bulls. As expected, adding both additional sources of information was the best strategy.

Updating the reference population in time is very important to keep the accuracy at a constant level across generations. In the context of performing GS for a hard to measure trait, relying on a cow reference population, Pszczola and Calus (2016) investigate the required number of cows that should be added every generation to keep the accuracy constant. When starting with 2000 cows in the initial reference population, on average 600 cows needed to be added per generation to achieve this. Of course GS is also promising for other livestock species than dairy cattle, although the economic return on GS may be more limited, because the relative benefit compared with pedigree-based selection is smaller, and therefore does not justify the costs to compile a large reference population. Shumbusho et al. (2016)

investigated the potential of GS in the meat sheep industry, considering different combinations of source of information used to select males. Assuming that an initial reference population was already available, carefully chosen GS strategies were able to yield up to 15% higher response than traditional selection, despite the higher costs of the selection process. Lillehammer *et al.* (2016) investigated the impact of GS in a pig breeding scheme for a trait not measured on the nucleus animals. Across several considered strategies, the genetic gain was not increased with GS if the trait had an economic value of 10% of the breeding goal, but was increased if this was 30%. In conclusion, the studies published in this issue collectively show that for scenarios where the size of reference populations is limited, there are still several solutions to achieve acceptable levels of accuracy, and thereby provide opportunities to improve accuracy of GS in small populations in practice.

References

Koivula M, Strandén I, Aamand GP and Mäntysaari EA 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal 10, 1061–1066.

Lillehammer M, Sonesson AK and Meuwissen THE 2016. Use of field data in pig genomic selection schemes – a simulation study. Animal 10, 1025–1032.

Lund MS, van den Berg I, Ma P, Brøndum RF and Su G 2016. How to improve genomic predictions in small dairy cattle populations. Animal 10, 1042–1049.

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

Pszczola M and Calus MPL 2016. Updating the reference population to achieve constant genomic prediction reliability across generations. Animal 10, 1018–1024.

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

Schöpke K and Swalve HH 2016. Review: opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal 10, 1050–1060.

Shumbusho F, Raoul J, Astruc JM, Palhiere I, Lemarié S, Fugeray-Scarbel A and Elsen JM 2016. Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. Animal 10, 1033–1041.

Su G, Ma P, Nielsen US, Aamand GP, Wiggans G, Guldbrandtsen B and Lund MS 2016. Sharing reference data and including cows in the reference population improve genomic predictions in Danish Jersey. Animal 10, 1067–1075.

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.