Quantitative methods in genomics and animal breeding

Mario Calus[1], Yvonne Wientjes[1,2], Roel Veerkamp[1]

[1]Animal Breeding and Genomics Centre, Wageningen Livestock Research, P.O. Box 135, 6700 AC Wageningen, The Netherlands

[2]Animal Breeding and Genomics Centre, Wageningen University, 6700 AH Wageningen, The Netherlands

Corresponding author: Mario Calus, mario.calus@wur.nl

Animal Breeding and Genomics Centre

Wageningen UR Livestock Research

P.O. Box 135

6700 AC Wageningen

The Netherlands

Phone: 0031-320-238265

Genomic selection

Genomic selection is rapidly being adopted in animal breeding (Hayes et al., 2009b; Calus, 2010; de los Campos et al., 2013). Genomic selection is an attractive alternative to traditional selection strategies, because it allows to increase genetic gain by increasing the intensity of selection, shortening the generation interval, and it may even increase the accuracy of selection, depending on the traits considered. All these characteristics jointly lead to an increase of genetic gain per generation (Meuwissen et al., 2013). Genomic selection relies on genomic breeding values, that can be predicted with relatively high accuracy at a very early age of the selection candidates. Genomic breeding values are estimated from a reference population (RP) with individuals with known phenotypes and SNP genotypes (Meuwissen et al., 2001). Effects of SNP genotypes are calibrated, by associating SNP genotypes to observed phenotypes or breeding values.

Genomic selection was initially presented to rely on linkage disequilibrium (LD) between SNPs and QTL. Early additional research on genomic selection, however, indicated that within-breed genomic selection is largely driven by genetic relationships that are captured by the SNPs (i.e. genomic relationships) between the RP and the selection candidates (Habier et al., 2007). In fact, it has been shown that breeding values of individuals with a high relationship to the RP, can be predicted with higher accuracy than breeding values of individuals with poor relationships to the RP (Habier et al., 2007; Habier et al., 2010; Clark et al., 2012; Pszczola et al., 2012a). This leads to another straightforward interpretation of the principle behind genomic selection, which is that the SNP information predominantly helps to explain the Mendelian sampling term of an animal (Daetwyler et al., 2007). SNP based relationships are much better in capturing the variance in true relationships than pedigree based relationships (Figure 1), because they capture the Mendelian sampling term.

Genomic prediction models

Many different models have been proposed to estimate genomic breeding values (for a review see: de los Campos et al. (2013)). A well-known categorization of genomic prediction models is by dividing them into a category that involves explicit estimation of SNP-effects, while another category involves using genomic relationships. Although those categories coincide with the alternative explanations of the principle of genomic selection, i.e. relying on SNP-QTL LD versus use of genomic relationships, it has been shown that some models from either category are in fact equivalent (Goddard, 2009). This stresses that the accuracy of genomic prediction cannot easily be separated into a component due to SNP-QTL LD and a component due to prediction of relationships between selection candidates and the RP. In fact, close relationships cause LD across long distance on the genome, and break down as a function of the relationships.

All of the proposed genomic prediction models have features that enable simultaneous estimation of a large number of SNP effects (p) based on a relatively small number of animals (n). Clear relationships exist between different genomic prediction models in their strategy to tackle this $n<<p$ problem (de los Campos et al., 2013). One important strategy is using a shrinkage estimation procedure. The random regression-BLUP (RR-BLUP) model (e.g. Habier et al., 2007), which uses the same variance for each SNP, is expected to show more shrinkage than variable selection methods, which are able to adapt the variance for each SNP, conditional on its estimated effect. This is illustrated in Figure 2, where estimated SNP-effects from RR-BLUP are plotted against estimated SNP-effects of BayesC, which is a variable selection method. This figure shows that the SNP that receive the largest effect in BayesC, also receive the largest effect in RR-BLUP. At the same time, the largest SNP effects

estimated with BayesC are substantially larger than those estimated with RR-BLUP, due to its flexibility by performing variable selection. In other words, SNP-effects estimated with RR-BLUP are more strongly affected by shrinkage than SNP-effects estimated with BayesC.

Multi-trait genomic prediction

Most research around genomic prediction focusses on single trait models. In traditional breeding programs, use of predictor traits in multi-trait breeding value estimation proved to be very successful in increasing the accuracy of selection. This suggests that the accuracy of genomic selection may also benefit from using predictor traits in multi-trait genomic prediction. So far, 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; Hayashi and Iwata, 2013).

The benefit of using multi-trait genomic prediction for US Holsteins was investigated in two studies. Using multi-trait genomic prediction for conception rate in the first three parities resulted in a doubling of the reliability (prediction accuracy squared) of genomic predictions compared to pedigree based 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 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 predictor traits (Pszczola et al., 2012b). This study showed that indeed the predictor traits improved the accuracy of prediction for dry matter intake, but 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 predictor traits, as it has been shown previously that the added benefit of genomic information in multi-trait models is decreased when predictor traits with strong genetic correlations are used (Calus and Veerkamp, 2011).

An important question is whether the increase in accuracy of multi-trait compared to single trait genomic prediction arises simply because SNP-effects are estimated with higher accuracy, or because the multi-trait genomic prediction model is better able to detect QTL compared to its single-trait counterpart. We have developed a bivariate Bayesian Stochastic Search Variable Selection (BSSVS) model that can use data from two traits that are each measured on a separate group of animals (Calus et al., 2013). This model was applied to a scenario where the one group of animals was a cow RP and the other group of animals was a bull RP. 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. Additionally, using the cow and bull data combined, resulted in increased power to detect QTL. Although some of the additional QTL may have been false positives, several of the additional QTL appear to be ‘true’ because they could be validated based on existing literature. In another study, evidence for two QTL related to progesterone levels, detected using the single-trait BSSVS model, considerably improved when the bivariate BSSVS model was applied using information on correlated fertility traits (Berry et al., 2012). These examples indicate that, at least to some extent, the increase in

accuracy of multi-trait compared to single trait genomic prediction is the result of increased power to detect QTL

Size of the reference population (RP)

The $n<<p$ problem encountered in genomic prediction models may be alleviated somewhat in the near future, because the number of animals in the RP are increasing rapidly due to several reasons. Firstly, the costs of genotyping are continuously decreasing. Secondly, depending on the species and the structure of the breeding program, large numbers of selection candidates are genotyped to increase the intensity of selection. Eventually, part of those selection candidates are promoted to become breeding animals. If those animals, their offspring or other close relatives, are phenotyped later on, then those reference candidates can be added to the RP. Thirdly, especially in dairy cattle, international exchange of genotypes is becoming common practice (e.g. Lund et al., 2011), because of the recognized mutual benefits. Also, now most dairy bulls have been genotyped, more and more cows are being genotyped, as well.

An interesting observation is that with an increase in the size of the RP, the differences in observed accuracies between models appear to decrease. This is illustrated in Figure 3, where accuracies of genomic prediction for fat and protein percentage in dairy cattle obtained from GBLUP and two Bayesian models are presented. The Bayesian models involved one model with marker specific shrinkage (BayesA), and a variable selection model (Bayes SSVS). The difference between GBLUP and the Bayesian models are manifested in the prior settings of the models. The explanation for the decrease in accuracy between models, when the number of animals in the RP increases, is that in Bayesian learning, with an increase in data size, the dependency on the priors of the model decreases.

50k SNP chips

The number of animals used in genomic prediction models of reported studies ranges from a few hundred, up to over 20,000 (e.g. Lund et al., 2011; Tsuruta et al., 2013). The number of SNPs used is typically ~40k. This results from the fact that chips with 50-60k are available for all major livestock species (Van Tassell et al., 2008; Matukumalli et al., 2009; Ramos et al., 2009; Groenen et al., 2011), of which generally 10-15k are removed during the editing process. Those 50k SNP chips are currently used as a standard genotyping platform for genomic selection. At the time when those 50k SNP chips were introduced, front-runners in dairy cattle breeding had already started to use genomic selection with panels of a few thousand SNPs (De Roos et al., 2009b). Using those few thousand SNPs, genomic prediction models simultaneously used LD and linkage information. At the time, an important question was whether linkage information still needed to be included in genomic prediction models when using the 50k SNP chip instead of only a few thousand SNPs. It was shown in a simulation study that LD alone captured with 50k SNP chips, capturing an average r^2 value between SNP > 0.2, indeed was sufficient to explain variation at the QTL, such that linkage analysis information did not need to be included in genomic prediction models anymore (Calus et al., 2008). Recently, it has been proposed that modelling linkage information may, however, be useful when genomic predictions are used across multiple generations (Habier et al., 2013).

Higher SNP density & whole genome sequence data

Although a rapid increase in the size of RPs is currently observed, the number of SNPs used is expected to increase rapidly as well. The main advantage of using a higher SNP density, is

that the SNPs used are physically closer to the QTL, which is expected to result in higher QTL-SNP LD and therefore in a higher persistency of accuracy of prediction across generations. For cattle, two higher density SNP chips have been developed, the Affymetrix Axiom Genome-Wide BOS1 Array (648,874 SNP) and the Illumina High-Density Bovine Bead Chip Array (777,962 SNP) (Rincon et al., 2011). For chicken, recently a higher density chip with 580,954 SNPs is developed (Kranis et al., 2013). For other species, thus far no higher density SNP chips are available.

Currently, early investigations are undertaken to use whole genome sequence data in genomic prediction (Meuwissen and Goddard, 2010; Hayes et al., 2012). In such applications, where perhaps >10,000,000 SNP are included, it has been suggested that variable selection methods such as BayesB are required to make optimal use of whole genome sequence data in genomic prediction (Meuwissen and Goddard, 2010).

Genomic prediction models for across-breed genomic prediction

As discussed previously, the accuracy of the genomic breeding value of a selection candidate heavily depends on its relationship with the RP. Following up on this idea, Wientjes et al. (2013) investigated the accuracy of genomic breeding values for selection candidates that were increasingly more related to the RP. This was achieved by simulating selection candidates that had genotypes sharing the following properties with the genotypes of the RP: 1) allele frequency, 2) LD structure, 3) haplotypes (837 across the genome), 4) haploid chromosomes, or 5) family structure. Those scenarios reflect increasing relationships with the RP, where the 5th scenario represents that RP and selection candidates originate from the same breed. Predicted accuracies showed that, especially for small RPs, family relationships between RP and selection candidates are very important (Figure 4). Predictions across

different sizes of RP indicated that for the scenario where selection candidates were related through short haplotypes to the RP, the RP had to be ~15 times as large to reach a reliability (squared correlation) of 0.6 compared to a scenario with family relationships between RP and selection candidates.

These results are in agreement with results in the literature on across-breed genomic prediction using 50k genotypes, that show limited or no increase in accuracy due to use of multi-breed RPs compared to single-breed RPs (Hayes et al., 2009a; Pryce et al., 2011). Previously, it has been predicted that in order to successfully combine RP of Holstein-Friesian and Jersey, at least 300,000 SNPs should be used (De Roos et al., 2008; de Roos et al., 2009a). Using > 600,000 SNPs on the Illumina High-Density Bovine Bead Chip Array, however, thus far only very slight increases in accuracy have been reported (Erbe et al., 2012).

Genomic prediction models – ungenotyped animals

The category of genomic prediction models that are based on genomic relationships, rather than explicit estimation of SNP-effects, are straightforward to implement, since it only involves replacing a pedigree based by a genomic relationship matrix. Several software packages are able to read in an external relationship matrix (e.g. Gilmour et al., 2009; Mulder et al., 2010). In addition, it has been shown that a genomic relationship matrix from a subset of genotyped animals, can conveniently be blended with a pedigree based relationship matrix of all animals in the population, and used in a so-called single-step approach (Aguilar et al., 2010; Christensen and Lund, 2010). This approach is termed “single-step”, because it computes breeding values in one step, simultaneously using information of animals with and without phenotypes. This provides an attractive alternative, to a two-step approach where two

sets of breeding values are computed separately using only genotyped animals or all animals simultaneously, and combined afterwards. Important challenges for the single-step approach, involve proper scaling of the pedigree-based and genomic relationship matrices, to make them compatible (Forni et al., 2011; Vitezica et al., 2011).

CONCLUSION

Genomic selection is revolutionizing breeding programs worldwide. Reference populations are growing rapidly, and are mainly genotyped using the common 50k SNP chips. To enable genomic selection, genomic breeding values are computed using various genomic prediction models. A large research effort is currently geared towards developing and testing such genomic prediction models. Genomic prediction models can be divided in models that explicitly estimate SNP effects and models that compute and use genomic relationships. Those two categories appear to coincide with two alternative explanations of the principle behind genomic selection. The first explanation is that genomic selection relies on SNP-QTL LD, and the second explanation is that genomic selection is driven by (genomic) relationships that are captured using the SNP. Nevertheless, both explanations largely explain the same mechanism at a different level, e.g. LD within a group of closely related animals is higher than in a group of less related animals. Future challenges include dealing with increasing numbers of animals with genotypes, as well as increasing numbers of genotypes per animal, up to >10,000,000 when using whole genome sequence data. An important unanswered question is whether this high SNP density will enable accurate use of information across breeds.

ACKNOWLEDGEMENT

The authors acknowledge financial support from the Dutch Ministry of Economic Affairs, Agriculture, and Innovation (Public-private partnership “Breed4Food” code KB-12-006.03-005).

REFERENCES

Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 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.

Aguilar, I., I. Misztal, S. Tsuruta, G. R. Wiggans, and T. J. Lawlor. 2011. Multiple trait genomic evaluation of conception rate in Holsteins. J Dairy Sci. 94:2621-2624.

Berry, D. P., J. W. M. Bastiaansen, R. F. Veerkamp, S. Wijga, E. Wall, B. Berglund, and M. P. L. Calus. 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.

Calus, M. P. L. 2010. Genomic breeding value prediction: methods and procedures. Animal. 4:157-164.

Calus, M. P. L., Y. de Haas, and R. F. Veerkamp. 2013. Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies.

Calus, M. P. L., T. H. E. Meuwissen, A. P. W. De Roos, and R. F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics. 178:553-561.

Calus, M. P. L., H. A. Mulder, and J. W. M. Bastiaansen. 2011. Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genet. Sel. Evol. 43:34.

Calus, M. P. L. and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43:26.

Christensen, O. F. and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet Sel Evol. 42.

Clark, S., J. Hickey, H. Daetwyler, and J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet Sel Evol. 44:4.

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124:369-376.

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 193:327-345.

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

De Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics. 179:1503-1512.

de Roos, A. P. W., C. Schrooten, and T. Druet. 2011. Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. J Dairy Sci. 94:4708-4714.

De Roos, A. P. W., C. Schrooten, E. Mullaart, S. Van der Beek, G. De Jong, and W. Voskamp. 2009b. Genomic selection at CRV. Pages 47-50 *in* Interbull Bulletin no. 39. Proc. of the Interbull International Workshop - Genomic Information in Genetic Evaluations, Uppsala, Sweden.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 95:4114-4129.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet Sel Evol. 43:1.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK.

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

Groenen, M., H.-J. Megens, Y. Zare, W. Warren, L. Hillier, R. Crooijmans, A. Vereijken, R. Okimoto, W. Muir, and H. Cheng. 2011. The development and characterization of a 60K SNP chip for chicken. 12:274.

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

Habier, D., R. L. Fernando, and D. J. Garrick. 2013. Genomic-BLUP Decoded: A Look into the Black Box of Genomic Prediction. Genetics. DOI 10.1534/genetics.113.152207.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5.

Hayashi, T. and H. Iwata. 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics. 14.

Hayes, B., C. Anderson, H. D. Daetwyler, R. Fries, B. Guldbrandtsen, M. Lund, D. Boichard, P. Stothard, R. F. Veerkamp, I. Hulsegge, D. Rocha, C. P. Van Tassel, D. Coote, and M. Goddard. 2012. Towards genomic prediction from genome sequence data and the 1000

bull genomes project. Page 55 *in* Book of Abstracts of 4th International Conference on Quantitative Genetics, Edinburgh.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. 41.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009b. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433-443.

Hickey, J. M. and G. Gorjanc. 2012. Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. G3: Genes, Genomes, Genetics. 2:425-427.

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

Kranis, A., et al. 2013. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics. 14:59.

Lund, M., S. de Roos, A. de 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. 43:43.

Matukumalli, L., C. Lawley, R. Schnabel, J. Taylor, M. Allan, M. Heaton, J. O'Connell, S. Moore, T. Smith, T. Sonstegard, and C. Van Tassell. 2009. Development and characterization of a high-density SNP genotyping assay for cattle. PLoS ONE. 4:e5350.

Meuwissen, T. H. E. and M. E. Goddard. 2010. Accurate prediction of genetic value for complex traits by whole-genome resequencing. Genetics. 185:623 - 631.

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

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2013. Accelerating improvement of livestock with genomic selection. Ann Rev Anim Biosci. 1:221-237.

Mulder, H. A., M. Lidauer, I. Stranden, E. A. Mantysaari, M. H. Pool, and R. F. Veerkamp. 2010. MiXBLUP manual. Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad, the Netherlands.

Pryce, J. E., B. Gredler, S. Bolormaa, P. J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling, J. Solkner, M. E. Goddard, and B. J. Hayes. 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. J. Dairy Sci. 94:2625-2630.

Pszczola, M., T. Strabel, H. A. Mulder, and M. P. L. Calus. 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., R. F. Veerkamp, Y. De Haas, T. Strabel, and M. P. L. Calus. 2012b. Predictor traits improve accuracy of genomic breeding values for scarcely recorded traits. Page 151 *in* Book of Abstracts of 4th International Conference on Quantitative Genetics, Edinburgh.

Ramos, A. M., et al. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS ONE. 4:8.

Rincon, G., K. L. Weber, A. L. Van Eenennaam, B. L. Golden, and J. F. Medrano. 2011. Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. J Dairy Sci. 94:6116-6121.

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

Tsuruta, S., I. Misztal, and T. J. Lawlor. 2013. Short communication: Genomic evaluations of final score for US Holsteins benefit from the inclusion of genotypes on cows. J Dairy Sci. 96:3332-3335.

Van Tassell, C. P., T. P. L. Smith, L. K. Matukumalli, J. F. Taylor, R. D. Schnabel, C. T. Lawley, C. D. Haudenschild, S. S. Moore, W. C. Warren, and T. S. Sonstegard. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Meth. 5:247-252.

Verbyla, K. L., B. J. Hayes, P. J. Bowman, and M. E. Goddard. 2009. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. Genet. Res. 91:307-311.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet Res. 93:357-366.

Wientjes, Y. C. J., R. F. Veerkamp, and M. P. L. Calus. 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics. 193:621-631.

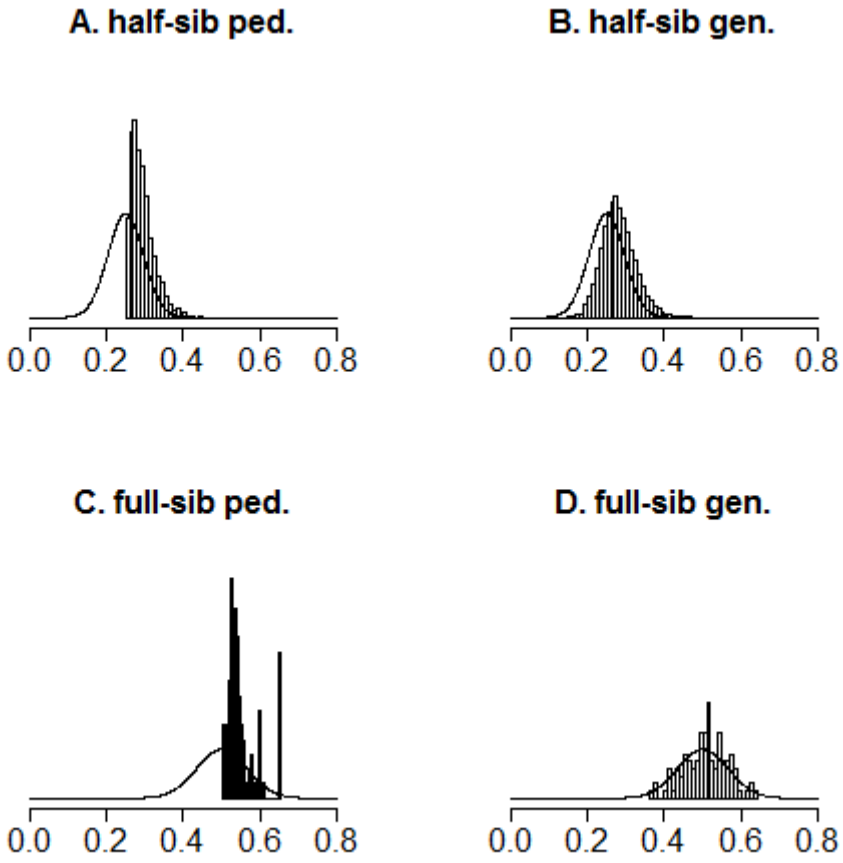

Figure 1. Empirical (bars) and expected distributions (smoothed lines) of half-sib (A & B) and fullsib relationships (C & D). Empirical distributions are based on pedigree (A & C) or genomic information (B & D). Source: Calus et al. (2011).

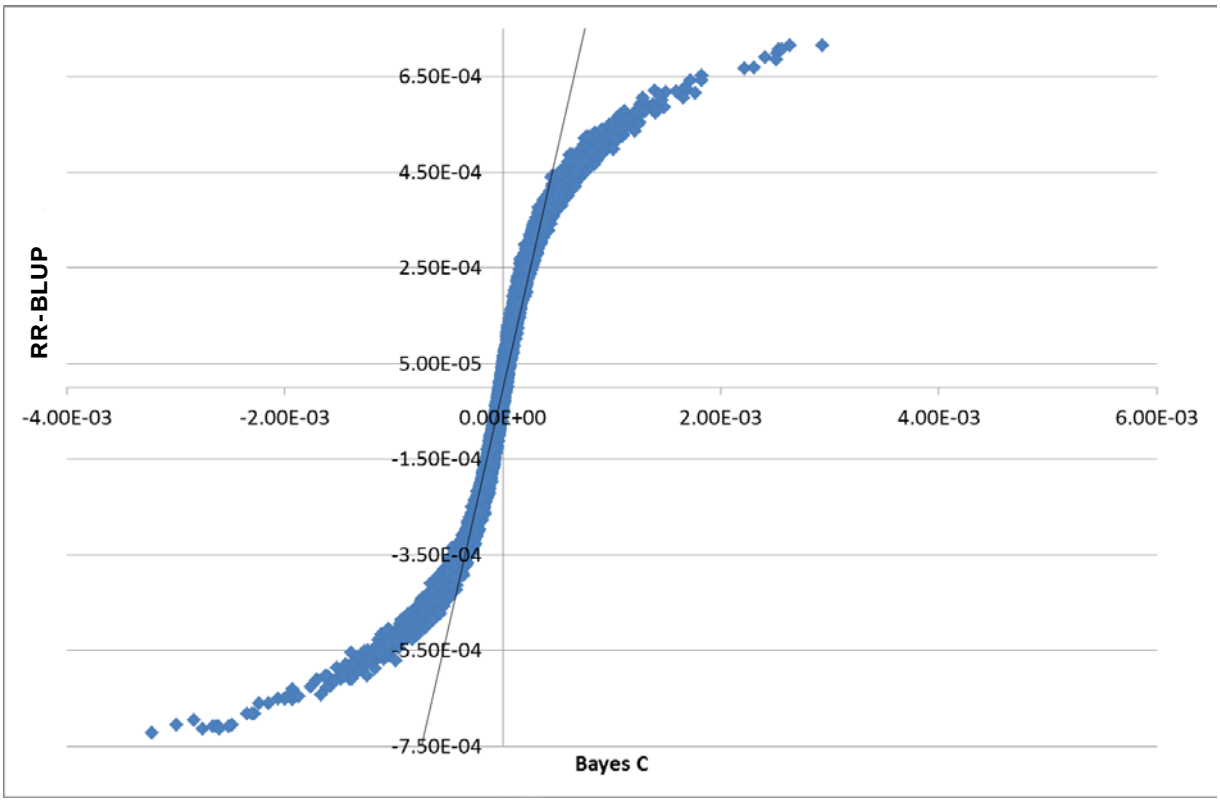

Figure 2. Estimated SNP (allele substitution) effects using BayesC versus RR-BLUP (unpublished results). The SNP-effects were estimated from one replicate of the simulated data described by Hickey and Gorjanc (2012).

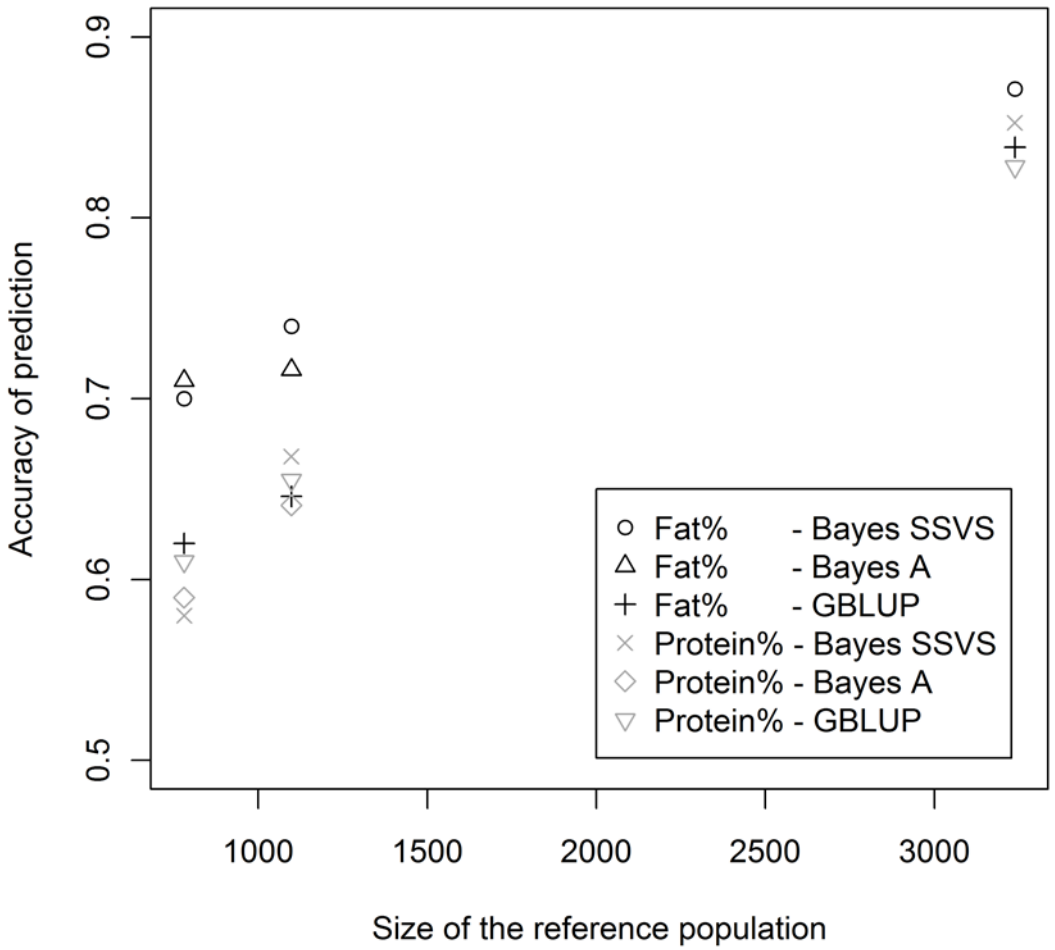

Figure 3. Accuracies of predictions from G-BLUP, BayesA and Bayes SSVS models for fat and protein percentage, estimated using Holstein-Friesian reference populations with different sizes. Source: de los Campos et al. (2013), based on results published elsewhere (Hayes et al., 2009b; Verbyla et al., 2009; de Roos et al., 2011).

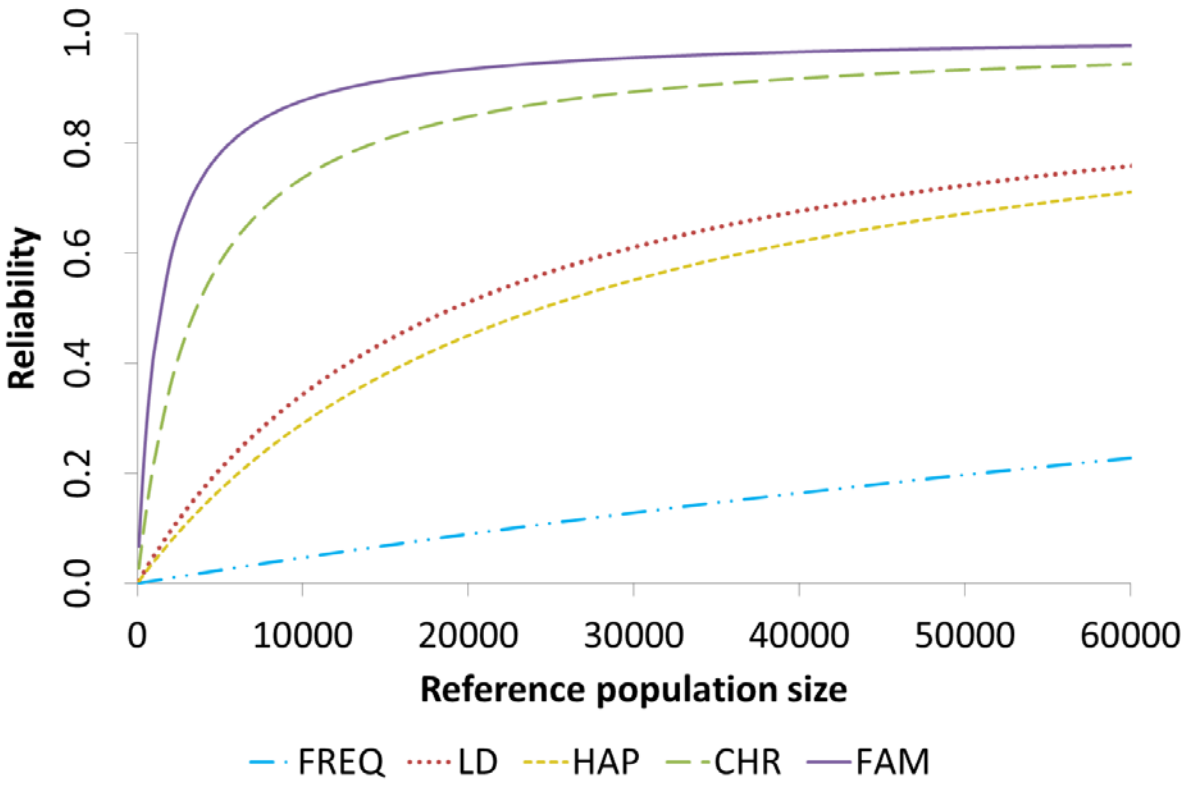

Figure 4. Predicted reliability of genomic prediction, at a heritability of 0.6 and different sizes of the reference population, obtained with the deterministic formula of Daetwyler et al. (2008) for the five different scenarios using different information sources from the reference population (from bottom to top). Selection candidates were simulated based on the following information of the reference population: allele frequency (FREQ), 837 haplotypes of equal length (HAP), LD pattern (LD), haploid chromosomes (CHR), and individuals from the reference population (FAM). Source: Wientjes et al. (2013).