Genetic variance and covariance components for across population evaluation of Brown Swiss cattle

Anne Loberg
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Licentiate thesis
Swedish University of Agricultural Sciences
Uppsala 2015

Institutionen för husdjursgenetik **Rapport 149**
Publication No. 149

Swedish University of Agricultural Sciences **Uppsala 2015**
Department of Animal Breeding ISSN 1401-7520
and Genetics ISRN
SLU-HGEN-R--149--SE

Cover picture: Anne Loberg

ISBN (print version) 978-91-576-9303-7
ISBN (electronic version) 978-91-576-9304-4

Print: SLU Repro, Uppsala 2015

Genetic variance and covariance components for across population evaluation of Brown Swiss cattle

Abstract

Dairy cattle breeding is an international business, with trade of animal material across populations. Interbull performs international genetic evaluations of dairy bulls, enabling fair comparisons across populations. The bull daughter performances differ between populations and international evaluations therefore require estimation of genetic correlations between the populations. Prerequisites for estimating correlations are knowledge about the genetic variances and covariances within and between populations and the relationship between the bulls. Traditionally, pedigree information has been used to build the relationship matrix connecting included bulls. The recent developments in DNA technology have made it possible to also build genomic relationship matrices using information from the bull genomes.

The purpose of this thesis was to evaluate variances and covariances estimated using pedigree or genomic relationship matrices and the effect of factors such as trait heritability, population size, and number of bulls used in more than one population. Moreover, we evaluate the genetic correlations estimated using our different estimates.

Genetic variance estimates were compared for a total of 175 population-trait combinations, and genetic covariance and correlation estimates were assessed for four populations and three traits. Phenotypes, genotypes and pedigree were available for 8 864 Brown Swiss bulls originating from Germany-Austria, France, Italy, Slovenia, Switzerland and the United States of America.

The results showed that more genetic variance and covariance was explained when the pedigree relationship matrix was used than using the genomic relationship matrix. We found an effect of trait heritability and population size on the genetic variance estimates and an effect of number of common bulls on the genetic covariance estimates. The estimated genetic correlations were similar for both relationship matrices used in the estimation of variances and covariances, indicating that it should be possible to obtain accurate across-population correlation estimates using genomic relationship matrices.

Keywords: Brown Swiss cattle, genetic variance, genetic covariance, genomic relationship matrix, across population genetic correlation

Author's address: Anne Loberg, SLU, Department of Animal Breeding and Genetics, P.O. Box 7023, 750 07 Uppsala, Sweden
E-mail: Anne.Loberg@ slu.se

It always seems impossible until it is done

Nelson Mandela

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Loberg, A., Dürr, J.W., Fikse, W.F., Jorjani, H., and Crooks, L. (2015). Estimates of genetic variance and variance of predicted genetic merits using pedigree or genomic relationship matrices in six Brown Swiss cattle populations for different traits. *Journal of Animal Breeding and Genetics, in press*

II Loberg, A., Crooks, L., Fikse, W.F., Strandberg, E., and Jorjani, H. Estimation of genetic covariance across four Brown Swiss cattle populations using pedigree or genomic relationship matrices (manuscript)

Paper I is reproduced with the permission of the publishers.

Abbreviations

A	Pedigree relationship matrix
BSW	Brown Swiss cattle
CHE	Switzerland
DEA	Germany-Austria
DGV	Direct genomic value
G	Genomic relationship matrix
GCTA	Genome-wide Complex Trait Analysis
GxE	Genotype by environment interaction
MACE	Multiple-across country evaluation
PGM	Predicted genetic merit
SNP	Single nucleotide polymorphism
UK	United Kingdom

1 Introduction

The Brown Swiss cattle (BSW) is a dairy breed originating from the Swiss Alps. The breed has several good characteristics, e.g., good longevity, adapting well to both warm and cold climates and strong feet and legs (Brown Swiss Association, 2015). In addition, the milk composition is favourable for cheese making (De Marchi *et al.,* 2008). Originally, BSW was bred to be a multi-purpose breed used for milk production, meat, and draught. The breeding goal has, however, changed over time to reflect changes in needs, and today an explicit dairy type of Brown Swiss cattle exists (Zogg, 1997). The breed is used in populations across the globe, with 10 populations participating in the Interbull international dairy evaluations.

The international evaluations are performed at the Interbull Centre with the Multiple Across Country Evaluation methodology (MACE) (Schaeffer, 1994), using a multiple–trait sire model. Service users participate in the international evaluation with national predicted genetic merits (PGM) and receive back international PGM in the population's own scale on all included bulls. The novelty of this method was that it treated the measured traits in different populations as non-identical but correlated traits. The correlations across populations are thus an indication of the amount of genotype by environment interaction (GxE) present between populations. The genotype by environment effects are caused by differences in estimation methodology, differences in average PGM between populations and differing environmental conditions (Philipsson, 1998). The inclusion of genetic correlations in the international evaluation enables fair ranking of bulls across populations. The size of the correlations varies depending on how similar populations and environments are. For example, correlations for Holstein and milk production between United Kingdom (UK) and the United States of America have been estimated at 0.86, between UK and Australia at 0.79 (Interbull, 2015) and between UK and Kenya at 0.49 (Ojango & Pollott, 2002). A prerequisite for estimating genetic correlations is knowledge about the genetic variances within and

genetic covariances between populations (Sigurdsson *et al.*,1996). In the international evaluations performed at the Interbull Centre, the genetic variances and covariances are estimated twice a year (Fikse, 2004).

Traditionally, the estimation of genetic variances and covariances has been done by including a pedigree relationship matrix (A) in the model. This matrix traces individuals back to the founder animals, which are assumed to be unrelated. The set of founder individuals is called the base population and the estimated genetic variances and covariances are for the base population. The pedigree used in the international evaluations is built by merging the population-specific pedigrees together. The estimation of genetic correlations across populations is made possible by having bulls with daughter records in more than one population, the so-called common bulls (Jorjani *et al.*, 2005; Sigurdsson *et al*., 1996).

With genomic information of animals now available it is also possible to build relationship matrices based on the genomic information. The genomic relationship matrix (G) is built using the genotypes at all measured loci. Using genomics should in theory enable us to explain more of the genetic variance present in populations and genetic covariance between populations, by identifying alleles that are identical by state even if they are not identical by descent according to the known pedigree. Also, the Mendelian sampling term can be captured in genomic relationship matrices. The potential value of this has been shown by comparing pedigree-based kinship with genomic-based kinship in pigs (Lopes *et al.*, 2013). By explaining more of the genetic variance, higher accuracy in genetic evaluations and an increased progress in breeding is expected (Henderson, 1984).

As the amount of bull genome information increased, several co-operations began to perform joint genomic evaluations (Loberg & Dürr, 2009). One of the co-operations was created in 2009 between the Interbull Centre and six organisations representing different Brown Swiss cattle populations (Santus, 2011). The main objective of these co-operations is to create the largest possible reference population and hence maximize accuracy (Hayes *et al.*, 2009). A reference population is a group of animals for which there are both genotypes and phenotypes available. This information can be used to estimate a value for each single nucleotide polymorphism (SNP), the allele effect. The sum of all estimated allele effects can be used to estimate genomic breeding values for animals without own phenotypes. This enables selection of young animals which shorten the generation interval, increase genetic gain and reduce the cost of proving bulls (Schaeffer, 2006).

The early expectation for genomic information was that more genetic variance would be explained than before. The first investigation of this using

genome wide association studies showed disappointing results. For instance, Aulchenko *et al.* (2009), showed the loci with significant effects for human height only explained a small fraction of the genetic variance estimated with traditional methods. Later, using 300 000 SNPs, Yang *et al.* (2010) explained 45% of the genetic variance in human height. In cattle, the amount of genetic variance explained by either pedigree or genomic relationship matrices has been described for many different populations and traits (Roman-Ponce *et al.*, 2014; Haile-Mariam *et al.*, 2013; Jensen *et al.*, 2012; Yang *et al.*, 2010), showing that the A matrix explains a larger amount of genetic variance than does the G matrix.

2 Aims of the thesis

The main aim of this thesis was to investigate the effect on the genetic correlation across populations of using different types of relationship matrices in the estimation. This was done by examining the components included in the genetic correlation estimation, the within-population genetic variances and the between-populations genetic covariance.
The more specific aims were to:

- Study how the genetic correlation across populations is affected by the use of different types of relationship matrices (G or A)
  - Investigate the impact of heritability and number of common bulls

- Quantify the proportion of genetic variance explained by a G matrix relative to the A matrix
  - Investigate the impact of trait heritability and population size

- Quantify the proportion of genetic covariance explained by a G matrix relative to the A matrix
  - Investigate the impact of heritability, number of common bulls and number of SNP-windows with large effects

- Investigate how the variance of PGM is affected by including more information in the prediction

3 Summary of investigations

3.1 Materials

3.1.1 Daughter performance information

National predicted genetic merits (PGM), genotypes and pedigree information for Brown Swiss bulls were provided from six populations: France, Germany-Austria, Italy, Slovenia, Switzerland and United States of America. A total of 34 different traits were available; for each trait between three and six populations were included, amounting to a total of 175 population-trait combinations. For Paper I all available population-trait combinations were included. For Paper II three traits were included, protein yield, stature and somatic cell score, for four populations: Germany-Austria, Italy, Switzerland and the United States of America.

3.1.2 Genotypic information

The genotypic information used in this thesis comes from an Interbull international genomic Brown Swiss evaluation (Uppsala, Sweden, April 2013). The single nucleotide polymorphisms (SNP) of 8 864 bulls were available, after imputation (VanRaden *et al.*, 2011) and edits 45 473 SNPs were left for analysis. In Paper I, information on all bulls was used and in Paper II information on 5 420 bulls was used.

In Paper II SNPs with large effects for the included traits were selected. The allele effects were grouped into SNP-windows of 10 sequential non-overlapping SNPs. For each SNP-window a mean allele effect was calculated, and a total mean and standard deviation across all SNP-windows were also calculated. Those SNP-windows with a mean deviating more than two standard deviations from the total mean were selected as having a large effect on the trait.

3.2 Methods

3.2.1 Relationship matrices

The pedigree relationship matrix (A) used in the estimations was the international sire-dam pedigree built at the Interbull Centre. The pedigree was built using the population-specific pedigrees submitted by all participating populations. The pedigree traces back to 1960 for each animal; animals born before 1960, animals with unknown sire and dam, and animals with only one progeny were treated as missing.

Two types of genomic relationship matrices (G) were used. Both were built using allele frequencies estimated in the international genomic evaluation of Brown Swiss cattle (Uppsala, Sweden, April 2013; VanRaden, 2008). The G matrix of VanRaden (2008) was used for the prediction of direct genomic values (DGV) in Paper I. The G matrix of Yang *et al*. (2011) was used in the estimation of the genetic variances in Paper I and covariances in Paper II.

3.2.2 Genetic analyses

In Paper I, genetic variances explained by the G matrix were estimated with Genome-wide Complex Trait Analysis (GCTA) using a mixed model with allele effects as random effects (Yang *et al.*, 2011), with international PGM as phenotypes. The genetic variances using an A matrix in the model were estimated in the international evaluation with MACE performed by the Interbull Centre (Uppsala, Sweden, 2013), with the national PGM as phenotypes. In Paper II, I estimated genetic variances and covariances with the DMU package using a bivariate mixed model (Madsen & Jensen, 2007) and the average information REML algorithm (Jensen *et al.*, 1997), using the national PGM as phenotypes and including the A matrix, G matrix or both in the estimation. For the estimation fitting both G and A in the model simultaneously, there were only results for protein yield and stature, due to convergence problems for somatic cell score. In the estimation, heritability within each population was kept constant and the residual covariances were assumed to be non-estimable. The assumptions behind these restrictions are that heritability estimated at the population level is the most accurate available and therefore the most suitable to use and PGM are estimated on different daughter groups in the included populations.

3.2.3 Comparison between estimates using G and A

Because of scale differences between populations, a comparison between methods using actual values becomes hard to interpret. All results for

estimations of variances and covariances are for this reason presented as ratios between estimates using G and/or A.

3.2.4 Statistical analysis of results

To calculate the means of national and international PGM and DGV the SAS software was used (SAS, 2013).

Linear models in SAS (SAS, 2013) were used to test for effects of heritability and population size on genetic variance ratios (Paper I). In Paper II, linear models were used to test for trait-specific effects of the number of common bulls and number of shared SNP-windows with large effects on genetic covariance ratios.

3.3 Main findings

3.3.1 Genetic correlation across populations

Genetic correlations were estimated using the genetic variances and covariances estimated within and between populations, using either the A matrix or a G matrix. The correlations estimated using the pedigree relationship was on average: 0.89 for protein, 0.96 for stature, and 0.85 for somatic cell score. The correlations estimated between populations using the genomic relationship matrix were very similar, differing at the most by 0.06 units. For protein yield and stature, the correlation when fitting both A and G was also estimated. Results were on average 0.01 lower than the correlation estimated with the A matrix.

A data set where no bulls have records in both considered populations was created for protein yield between Germany-Austria and Switzerland. This removes the link between the populations that is essential for across population genetic correlation estimations using the A matrix. This data set was used in variance component estimation using a G matrix (Paper II). The estimated between-population correlation without common bulls was 0.92, which is very close to the correlation of 0.93 estimated with the A matrix using a well-connected data set.

3.3.2 Genetic variance

Genetic variances estimated as part of the international evaluation using the A matrix were larger than the genetic variances estimated with GCTA using a G matrix (Paper I). The estimates using the G matrix explained 10-60% of the estimates when using the A matrix. The ratio between the two estimates increased with increased trait heritability. A small effect of population size was also seen, the ratio between estimates decreased with larger population sizes.

3.3.3 Genetic covariance

Genetic covariances estimated using the A matrix were always larger than the genetic covariances estimated using a G matrix, when these matrices were fitted separately. The G matrix explained on average 86% of the covariance explained by the A matrix. When the A and G matrices were fitted simultaneously in the model, the G matrix explained a larger part of the covariance than the A matrix, on average 80% of the total genetic covariance (Paper II).

The ratio between the covariance estimated using G or A matrices in the model decreased as the number of common bulls between populations rose, for stature; with more common bulls, a lower proportion of covariance was explained by the G matrix. But there was no significant relationship between the traits protein yield and somatic cell score with the number of common bulls. Similar results were seen for the ratio between estimates using G or the total covariance using G and A simultaneously in the model. There was a relationship with the number of common bulls and decreasing ratio for stature; this was not seen for protein yield.

The proportion of genetic covariance explained by the G matrix within the model fitting G and A simultaneously, showed contrasting results to the relationship between the number of common bulls and the ratios discussed above. There was a relationship for protein yield with number of common bulls, with a decreasing proportion of covariance explained by the G matrix with an increasing number of common bulls, but no relationship for stature.

The ratios discussed above were also tested for a relationship with the number of common SNP-windows and average heritability between traits, but no significant results, above the effect of the number of common bulls, were found.

3.3.4 SNPs with large effect

In Paper II, I studied the proportion of selected SNP-windows with large effects for a trait in two populations, and the proportion of selected SNP-windows that were in common for all populations within a trait. There were small differences in the number of common SNP-windows with large effects between populations within a trait. However, there were greater differences between traits. Stature had the largest proportion of SNP-windows with large effect in common between population pairs (77-86%) and in common for all tested populations (73%). For somatic cell score, the proportion in common between population pairs was 64-83% and common to all populations was 53%. Protein yield showed the smallest proportion of common SNP-windows between population pairs, 48-64%, and common to all populations, 31%.

3.3.5 Predicted genetic merits

In Paper I, the size of the variance for national PGM, international PGM and DGV was compared. Whilst the majority of our analyses otherwise were based on estimating genetic variances these are estimates on how far the genetic merits are spread out. Variances of PGM increased as more information was included in the estimation of PGM. The national PGM had the lowest estimated variance. These are estimated within each population using population-specific models and pedigrees. In the international evaluation, all national PGM and population-specific pedigrees are joined, creating new pedigree ties and adding new daughter information to the bulls that have daughters evaluated in more than one population. The variance of international PGM were on average 5% larger than the variance of national PGM. The genomic evaluations utilize animal information obtained from the international evaluation through the use of international PGM as phenotypes. In addition, the genomic information from the G matrix is included in the estimation of DGV. The variance of DGV was on average 22% larger than the variance of international PGM.

The ratio of variances of DGV to national or international PGM depended on trait heritability and population size. The increase in variance between the PGM and DGV was largest for low heritability traits and small populations.

4 General discussion

4.1 Genetic correlations and covariances across populations

Genetic correlations are estimated by dividing the between-population genetic covariance with the product of the square root of the genetic variance for each population. In Paper II, the genetic correlations across four populations were estimated for three traits: protein yield, stature and somatic cell score, using a model including either the A matrix, a G matrix or both. The difference between the estimated correlations for each population combination was small, on average 0.015 between A and G correlations and 0.01 between A and A + G correlations. When looking at the variance and covariance components included in the estimation of genetic correlations the relative size of estimates from a model including an A matrix was larger than when including a G matrix.

In Paper I and in several other studies (e.g. Haile-Mariam *et al.*, 2013; Jensen *et al*., 2012; Veerkamp *et al.*, 2011; Yang *et al*., 2010), models using the A matrix explained a larger amount of genetic variance than models using a G matrix. The ratio between estimates was found to be affected by trait heritability and population size (Paper I), number of markers included (Jensen *et al.*, 2012), reliability of the trait (Roman-Ponce *et al.*, 2014) and reference population size (Haile-Mariam *et al*., 2013).

For genetic covariance, the results in Paper II showed the same trend as for genetic variance – a larger amount of genetic covariance was explained when using a model with the A matrix than when using a G matrix. The A matrix also explained a larger amount of genetic covariance than when both A and G matrices were included in the model. This was unexpected, because I assumed that with inclusion of both types of relationship matrices more genetic covariance would be accounted for. The explanation could be that the model gave genetic covariance estimates for each relationship matrix, but there could

also be a covariance between them, which was not estimated and therefore not included in the estimated covariance.

There was a relationship between the number of common bulls between populations and the ratio between estimated genetic covariance using the G or A matrices, for stature. With an increasing number of common bulls the separation between the two types of estimates increased. This is probably caused by an increasing ability of the A matrix to capture the genetic covariance between populations with increasing pedigree ties between them. The proportion of genetic covariance explained by the G matrix in the model with A and G both included was also related to the number of common bulls, but the relationship was now significant only for protein yield. This suggests that the proportion of covariance explained by G or A when fitting them simultaneously can be affected by the number of common bulls between populations. This is interesting but with only two traits very inconclusive and more work is required to better understand this.

We tested whether more genetic covariance between populations could be explained by the G matrix relative to the A matrix when a higher proportion of SNPs with large effects were common to both populations. No significant effect was found but this could be due to a small number of data points.

The size of the estimated genetic correlations was tested against the factors found to affect genetic variance and covariance. To test if the heritability of the trait had an effect, the mean of the two population heritabilities was used. A significant effect was found, higher average heritability between included traits gave larger genetic correlation estimates, both using the A matrix and G matrix. This indicates that there is a stronger estimated genetic correlation when there is a larger genetic contribution to the phenotype.

Both components of genetic correlation across populations, genetic variance within populations and genetic covariance between populations, were affected by using different types of relationship matrices (A or G). However, the genetic correlation only showed small differences, indicating that even though the model including the G matrix explained less of the genetic variance and covariance, it can estimate the genetic correlation equally well as a model including the A matrix.

4.2 Genetic correlations without common bulls

A reduced data set was created between Switzerland (CHE) and Germany-Austria (DEA). For each bull with records in both populations, the record with the lowest EDC was removed. Using this data set it was possible to estimate genetic correlations without any common bulls, using the G matrix. The results

(Paper II) show only a 0.01 difference in genetic correlation compared with using the A matrix in the model with and without common bulls between the two populations. The estimates using the G matrix with or without common bulls were also very similar. This suggests that genetic correlations estimated with a G matrix may be independent of the number of common bulls. Hence, a G matrix could be useful for estimation of genetic correlations between populations where one or both have insufficient pedigree records.

4.3 Variance estimates using different relationship matrices

The difference in genetic variance estimated using the A matrix and a G matrix in Paper I, was larger than found in other cattle studies (Roman-Ponce *et al.*, 2014; Haile-Mariam *et al.*, 2013; Jensen *et al.*, 2012) and also larger than found in Paper II (not presented). In the bivariate analyses performed in Paper II using DMU, the genetic variances for included populations are available. Taking the average ratio between genetic variance estimated with the G and A matrix for each population across all estimations for each trait, the amount of genetic variance explained by the G matrix was 84-86% of the value obtained with the A matrix for protein, 77-95% for stature, and 83-91% for somatic cell score – these levels were more in line with the results from other cattle studies.

In both Paper I and Paper II, the A matrix was the international pedigree built at the Interbull Centre and the G matrix was built using the GCTA package and method by Yang *et al.* (2011). In Paper II all the estimations were done using DMU (Madsen & Jensen, 2007) using national PGM as phenotypes; the only difference was which relationship matrix was used in the estimation. In Paper I, on the other hand, the genetic variance estimates using the A matrix were those estimated in the regular international evaluation performed at the Interbull Centre, using national PGM as phenotypes and MACE. The genetic variances using the G matrix were estimated with international PGM as phenotypes using the estimation of the variance explained by the "all the SNPs" option in the GCTA package (Yang *et al.*, 2011). Hence, the genetic variance estimates in Paper I are influenced by not just the relationship matrix used but may also be affected by the different estimation procedures and phenotypes used.

4.4 Impact of amount of information in genetic evaluation

The variance of predicted genetic merits (PGM) increased with inclusion of more information in the evaluation: in Paper I it was shown that the variance among international PGM was larger than for national PGM. The variance

increased when the international pedigree was included in the evaluation; more relatives for included bulls are identified but most importantly more data are added in the evaluation (Mark *et al.,* 2002). When genomic information was added, the size of the variance increased further, the variance of direct genomic values (DGV) estimated with international PGM as phenotypes was on average 22% larger than the variance of international PGM. The increase in variance from PGM and DGV was largest for the traits with lowest heritability. This agrees well with the fact that traits with low heritability have larger portions of error variance and consequently more room for improvement. Larger variance among genetic merits reflects higher accuracies and thus better ranking of animals.

4.5 Differences in trait genetic architecture

The genetic correlation across populations describes how strongly the expression of a trait in one population is connected to the expression of the same trait in another population. This difference in expression is caused by GxE between populations (Philipsson, 1998). Most of the records in estimations of genetic correlations across populations are measured on bulls' daughters in only one population, only a limited number of bulls have daughter records in more than one population. In Paper II, the number of common bulls between populations consisted of between 5-10% of the total number of bulls included. However, no individual cow had records in more than one country.

One way to look at the nature of genetic correlation on the SNP level is to use the classification by Bohren *et al.* (1966). SNPs are grouped considering their effect on each of the included traits and if they have effects in the same or opposite direction. In Paper II it was shown that for each population pair, a large proportion (48-84%) of the SNP-windows with a large effect in one population, also had an effect in the other, and none of the SNP-windows showed opposite effects in the two populations. In contrast, van Binsbergen *et al.* (2012) investigating genetic correlation between milk production traits for Holstein cattle (e.g., milk yield and protein yield), found a much lower proportion of SNPs (0-17%) with large effects in two populations and they also found SNPs with effects in opposite directions in some cases. For the traits we examined in BSW, the SNPs showing the largest effect are similar across populations and any differences in selection across populations has not greatly affected which SNPs are most influential.

The difference between the genetic covariance estimated using the A matrix or a G matrix, and the relationship with number of common bulls varied across investigated traits (Paper II). This could be caused by trait architecture

differences, Hayes *et al*. (2010) showed that differences in the genetic architecture between traits can affect the results from statistical analyses. For example, they found that for the conformation trait Overall Type the accuracies of genomic PGM were similar when using a normal distribution or a heavy-tailed distribution. For fat percentage and proportion of black colour, two traits known to be influenced by genes with large effects (Hayes *et al.*, 2010; Grisart *et al.*, 2002), however, the heavy-tailed distribution gave higher accuracies.

5 Conclusions

Based on the results included in this thesis my conclusions are:

- Estimated genetic correlations across populations are not affected by the use of the A or G matrix
- More genetic variance within and genetic covariance between populations is explained when the A matrix rather than a G matrix is used.
- Trait heritability affects the relationship between variance estimates using A and G matrices. The G relationship matrix can explain more of the genetic variance explained by the A matrix with increasing heritability.
- The size of the populations affects the relationship between variance estimates using A and G relationship matrices. The G matrix can explain more of the genetic variance explained by the A matrix with increasing population size.
- For stature, the G matrix explains less of the genetic covariance explained by the A matrix as the number of common bulls increases.
- The variance among PGM is increased when a G matrix is used rather than the A matrix.

6 Future research

The current trend in animal breeding is to include more and more genomic information on the animals. This can be done using various different approaches. Independent of which approach will be the preferred one in the future, some research topics in the area of genetic correlations across populations would be interesting.

- It would be interesting to re-estimate genetic variances for all population-trait combinations, keeping everything but the relationship matrix constant, as in Paper II. Those results would not only give a measurement on the proportion of genetic variance explained by the G matrix relative to the A matrix. It would also be possible to investigate if the difference in methodology and phenotypes in the genetic variance estimation in Paper I, have a consistent impact on the genetic variance across trait estimates or not. This could be done by regressing the ratio between the new results using G and using A against heritability, the slope could then be compared to the one presented in Paper I.
- As a proof-of-principle, it was shown to be possible to estimate genetic correlation between two populations without any common bulls. It would be interesting to investigate if this could be a way of including populations with limited pedigree information in the international evaluation of bulls.
- In Paper II results showed differences between the included traits regarding their relationship with the number common bulls. It would be interesting to include more traits in similar analyses to investigate these trait differences further.

7 Genetiska varians- och kovarianskomponenter skattade mellan Brown Swiss populationer

7.1 Introduktion

Brown Swiss är en koras med ursprung i de schweiziska alperna. Den har lång livslängd, klarar sig bra i både varma och kalla klimat samt har starka ben. Dessa och andra positiva egenskaper hos rasen har gjort den populär över hela världen. I dagsläget deltar 10 populationer i de internationella avelsvärderingarna av mjölktjurar som utförs av Interbull Centre i Uppsala.

Den internationella avelsvärderingen utförs med metoden MACE (Multiple Across Country Evaluation MACE)(Schaeffer, 1994). Medlemsorganisationerna deltar i avelsvärderingen med nationella avelsvärden på de inhemska tjurarna. Efter avelsvärderingen får de tillbaka internationella avelsvärderingar, på den egna nationella skalan för alla inkluderade tjurar, det vill säga inte bara deras egna (Fikse, 2004). Miljö, skattningsmetoder samt genetisk nivå skiljer sig mellan populationer och bidrar till genotyp-miljösamspel mellan olika populationer eller annorlunda uttryckt att den genetiska korrelationen mellan populationer är lägre än 1.0 (Philipsson, 1998). Dessa genotyp-miljösamspel inkluderas i MACE och möjliggör en rättvis rangordning av inkluderade tjurar i olika miljöer. Den genetiska korrelationen påverkas av hur lika populationer är, som exempel har korrelationen mellan Storbritannien och USA, Australien samt Kenya för holsteinrasen och mjölkproduktion skattats till 0,86; 0,79 (Interbull, 2015) respektive 0,49 (Ojango & Pollot, 2002). Genetiska korrelationer skattas med hjälp av genetisk varians inom population samt genetisk kovarians mellan populationer och görs två gånger per år av Interbull Centre (Fikse, 2004).

I den traditionella internationella avelsvärderingen inkluderas en stamtavla som kopplar tjurarna till varandra. Den stamtavla som används är en

sammanslagning av populationsspecifika stamtavlor. För att finna släktband över populationsgränser och få en hög säkerhet på de internationella avelsvärdena krävs det att en del av tjurarna har avelsvärderingar (dvs producerande döttrar) i mer än en population (Jorjani *et al.*, 2005; Sigurdsson *et al.*, 1996).

Den tekniska utvecklingen inom molekylärgenetiken har lett till att en stor del av djurs och människors arvsmassa kan bli kartlagda till en överkomlig kostnad. Tillgång till denna nya så kallade genomiska information har lett till en drastisk ökning av metodutveckling och forskning. Ett av användningsområdena är att konstruera en släktskapsmatris byggd av informationen från tjurars arvsmassa. Teorin är att en släktskapsmatris byggd på genomisk information borde hitta mer genetisk variation mellan djuren än en stamtavla. Genom att finna en större del av den genetiska variansen kan säkerheten i avelsvärderingarna öka samt även avelsframsteget (Henderson, 1984).

I samband med att den genomiska informationen för tjurar blev tillgänglig i större skala började ett antal internationella genomiska avelsvärderingssamarbeten (Loberg & Dürr, 2009). Dessa samarbeten blev aktuella och nödvändiga eftersom det krävs genomisk information om ett stort antal djur för att uppnå samma säkerhet i de genomiska avelsvärderingarna som i de traditionella (Hayes *et al.*, 2009). Ett av dessa samarbeten bildades mellan Interbull Centre och sex Brown Swiss organisationer, Intergenomics (Santus, 2011).

Syftet med denna avhandling är att undersöka hur de genetiska korrelationerna mellan populationer påverkas av användandet av en genomisk släktskapsmatris istället för en stamtavla. Detta görs främst genom att undersöka den genetiska korrelationens komponenter genetisk varians och kovarians.

7.2 Sammanfattning av studierna

Två studier har inkluderats i detta arbete. I båda har avelsvärden, släktskapsinformation samt genomisk information från Brown Swiss tjurar använts. I den första studien inkluderades tjurar från sex olika länder och 34 egenskaper, totalt skattades genetisk varians för 175 land-egenskapskombinationer. Genetisk varians skattades i den internationella avelsvärderingen, samt med en genomisk metod. Den första studien inkluderade även en skattning av spridningen av tre olika typer av avelsvärden: nationella avelsvärden, internationella avelsvärden samt genomiska avelsvärden. Syftet var att jämföra spridningen mellan de olika avelsvärdena

och se hur den förändras när ny information om djuren tillkommer i skattningen.

I den andra studien inkluderades fyra länder och tre egenskaper: kilo protein, reslighet och celltal. Den genetiska kovariansen och den genetiska korrelationen skattades mellan länderna med olika släktskapsmatriser, stamtavla (A), genomisk släktskapsmatris (G) eller båda två inkluderade samtidigt.

Förhållandet mellan genetiska korrelationerna, varianserna och kovarianserna skattade med olika släktskapsmatriser har undersökts. Dels för att kunna avgöra hur väl den genomiska slaktskapsmatrisen fungerar i relation till den traditionella metoden att använda stamtavlan, dels för att undersöka om det finns faktorer som kan påverka storleken av skattningarna. De faktorer jag testade var arvbarhet, populationsstorlek, antalet gemensamma tjurar och antalet SNP:ar (DNA-markörer) med stor effekt.

Resultaten visar att den skattade genetiska korrelationens storlek inte påverkas av vilken släktskapsmatris som använts, men högre genomsnittlig arvbarhet mellan länder ger en högre genetisk korrelation. Storleks-förhållandet mellan de genetiska korrelationerna skattade med olika släktskapsmatriser jämfördes även med antalet gemensamma tjurar, men inget samband hittades.

Skillnaden mellan de genetiska varianserna skattade i den första studien var stor, de genomiskt skattade varianserna kunde förklara 10-60% av den genetiska varians som skattats i den internationella avelsvärderingen. Att skillnaden mellan de skattade varianserna var så hög kan förklaras med att det inte bara är släktskapsmatrisen som är olika utan även metod och fenotyper. Egenskapernas arvbarhet påverkade hur stor skillnaden är mellan de olika skattningarna, skillnaden minskade med ökad arvbarhet. Spridningen av avelsvärden ökade när mer information inkluderades i skattningen. Spridningen av de internationella avelsvärdena var i genomsnitt 5 % högre än för de nationella avelsvärdena och spridningen ökade ytterligare 22 % för de genomiska avelsvärdena.

Kovariansen skattad med G matris förklarade i snitt 85 % av den kovarians som skattades med A. För reslighet fanns det ett samband mellan antalet gemensamma tjurar och förhållandet mellan skattningarna. Med fler antal gemensamma tjurar förklarade G matrisen en mindre del av A matrisen. Detta samband kunde inte ses för kilo protein eller celltal. Andra faktorer, gemensamma SNP:ar och arvbarhet, som skulle kunna ha ett samband med hur stor del av kovarians som förklaras med G matrisen testades, men med det begränsade antal observationer kunde inga samband hittas.

7.3 Slutsatser i korthet

Den skattade genetiska korrelationen påverkas inte av vilken typ av släktskapsmatris som används vid skattningen av varianskomponenterna.

Högre genetisk varians och kovarians skattas när en stamtavla inkluderas i modellen än om en genomisk släktskapsmatris används. Skillnaderna mellan den skattade genetiska variansen påverkas av egenskapernas arvbarhet samt storleken på populationen. Den minsta skillnaden är för egenskaper med hög arvbarhet i stora populationer. Skillnaden i skattad kovarians påverkas av antalet tjurar de inkluderade populationerna har gemensamt. Med en ökad mängd gemensamma tjurar ökar skillnaden, för reslighet. Detta beror antagligen på att mer genetisk kovarians kan förklaras när den internationella stamtavlan får en ökad täthet.

8 References

Aulchenko, Y.S., Struchalin, M.V., Belonogova, N.M., Axenovich, T.I., Weedon, M.N., Hofman, A., Uitterlinden, A.G., Kayser, M., Oostra, B.A., van Duijn, C.M., Janssens, A. & Borodin, P.M. (2009). Predicting human height by Victorian and genomic methods. *European Journal of Human Genetics*, 17(8), pp. 1070-1075

Bohren, B.B., Hill, W.G. & Robertso.A (1966). SOME OBSERVATIONS ON ASYMMETRICAL CORRELATED RESPONSES TO SELECTION. *Genetical Research*, 7(1), pp. 44

Brown Swiss Association (2015) (http://www.brownswissusa.com/Breed/BrownSwissBreed/tabid/173/Default.aspx; last accessed 06 05 2015)

De Marchi, M., Bittante, G., Dal Zotto, R., Dalvit, C. & Cassandro, M. (2008). Effect of Holstein Friesian and Brown Swiss breeds on quality of milk and cheese. *Journal of Dairy Science*, 91(10), pp. 4092-4102.

Fikse, F.W. (2004) INTERBULL guides through the labyrinth of national genetic evaluations. In: *Proceedings of 55th meeting of the EAAP*, Bled, Slovenia.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. & Snell, Russel. (2002). Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine GDAT1 gene with major effect on milk yiled and composition. *Genome Research.* 12. pp. 222-231

Haile-Mariam, M., Nieuwhof, G.J., Beard, K.T., Konstatinov, K.V. & Hayes, B.J. (2013). Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. *Journal of Animal Breeding and Genetics*, 130(1), pp. 20-31.

Hayes, B.J., Pryce, J., Chamberlain, A.J., Bowman, P.J & Goddard, M.E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk.fat percentage, and type in Holstein cattle as contrasting model traits. *PLOS Genetics*, 6, e1001139

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), pp. 433-443.

Henderson, C.R. (1984). *Applications of linear models in animal breeding*. University of Guelph Press, Guelph, Canada

Interbull centre (2015) (http://www.interbull.org/ib/maceev_archive; last accessed 05 05 2015)

Jensen, J., Su, G. & Madsen, P. (2012). Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. *BMC Genetics*, 13.

Jensen, J., Mäntysaari, E.A., Madsen, P. & Thompson, R. (1997). Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *J.Ind. Soc. Agric. Stat.* 49, pp. 215-236.

Jorjani, H., Emanuelson, U. & Fikse, W.F. (2005). Data Subsetting Strategies for Estimation of Across-Country Genetic Correlations. *Journal of Dairy Science*, 88(3), pp. 1214-1224.

Loberg, A. & Dürr, J.W. (2009). Interbull survey on the use of genomic information. In: *Proceedings of Intbull workshop*, Uppsala, pp. 3-14.

Lopes, M.S., Silva, F.F., Harlizius, B., Duijvesteijn, N., Lopes, P.S., Guimarães, S.E.F. & Knol, E.F. (2013). Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. *BMC Genetics*, 14.

Madsen, P. & Jensen, J. (2007). *A Users's guide Guide to DMU, A Package for Analyzing Multivariate Mixed Models.* (Version: 6 release 5.2) [Computer Program]. University of Aarhus, Faculty Agricultural Sciences, Dept. of Genetics and Biotechnology, Research Centre Foulum, Denmark.

Mark, T., Fikse, F.W., Emanuelson, U. & Philipsson, J. (2002). Short Communication: Effect of Phantom Parent Grouping and Properties of Deregression for a Low Heritability Trait. *Journal of Dairy Science*, 85, pp. 2393-2395.

Ojango, J.M.K. & Pollott, G.E. (2002). The relationship between Holstein bull breeding values for milk yield derived in both the UK and Kenya. *Livestock Production Science*, 74(1), pp. 1-12.

Philipsson, J. (1998). Global use of bulls and the "INTERBULL system". *Acta Agriculturae Scandinavica - Section A: Animal Science*, 29(SUPPL.), pp. 98-107.

Roman-Ponce, S.-I., Samore, A.B., Dolezal, M.A., Bagnato, A. & Meuwissen, T.H. (2014). Estimates of missing heritability for complex traits in Brown Swiss cattle. *Genetics, selection, evolution : GSE*, 46(1), pp. 36-36.

Santus, E. (2011). Intergenomics: business rules and transition into service. In: *Proceedings of Interbull Workshop*, Guelph, Canada.

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123, pp 218-223

Schaeffer, L.R. (1994). Multiple-Country comparison of dairy sires. *Journal of Dairy Science*, 77, pp. 2671-2678.

Sigurdsson, A., Banos, G. & Philipsson, J. (1996). Estimations of Genetic (Co)variance Components for International Evaluation of Dairy Bulls. *Acta Agric. Scand., Sect. A, Animal Sci*, 46, pp. 129-136.

van Binsbergen, R., Veerkamp, R.F. & Calus, M.P.L. (2012). Makeup of the genetic correlation between milk production traits using genome-wide single nucleotide polymorphism information. *Journal of Dairy Science*, 95(4), pp. 2132-2143.

VanRaden, P.M., O'Connell, J.R., Wiggans, G.R & Weigel, K.A. (2011). Genomic evaluations with many more genotypes. *Genetics, Selection, Evolution.*,43, 10.

VanRaden, P.M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), pp. 4414-4423.

Veerkamp, R.F., Mulder, H.A., Thompson, R. & Calus, M.P.L. (2011). Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *Journal of Dairy Science*, 94(8), pp. 4189-4197.

Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*, 88(1), pp. 76-82.

Yang, J.A., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E. & Visscher, P.M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7), pp. 565-569.

Zogg, M. (1997). The Breeding Aim in the course of Time. In: Harder, M. & Grüter, O. (eds) *100 Years Swiss Brown Cattle Breeder's Federation.* Zug: Kündig Druck.

Acknowledgements

The work presented in this thesis was performed at the Department of Animal Breeding and Genetics at the Swedish University of Agricultural Sciences. I gratefully acknowledge financial support from Interbull. I am also thankful to the Intergenomics partners: Arbeitsgemeinschaft der Österreichischen BraunviehZüchter, Austria; Arbeitsgemeinschaft Deutsches Braunvieh, Germany; Associazione nazionale allovatori bovini della razza Bruna, Italy; Brown Swiss Association of the US, USA; Brune Genetique Services, France; Braunvieh Schweiz, Switzerland; and Zveza rejcev govedi rjave pasme Slovenije, Slovenia, for providing the data and taking an active interest in the project.

Many people have contributed to this thesis in different ways and I would like to express my sincere gratitude to all of you!

I would like to thank my advisors. My main advisor Erling Strandberg, thank you for your scientific guidance and your support, I truly appreciate all the small things. Lucy Crooks, thank you for everything you have been amazing. You have greatly improved the work in this thesis with your comments and suggestions and I will always remember that you believed in me.

I also want to thank all the people I have met over the years at the Interbull Centre, visitors and staffs, nothing is more inspiring than working in such multicultural environment. Some of you I would like to thank in person. Hossein Jorjani, thank you for everything you have taught me, I will always value our long discussions about science, genetics and life. João Dürr, thank you putting things in perspective and making me think in new ways. Freddy Fikse, thank you for giving me the opportunity to start working at the Interbull Centre and for everything you taught me about international evaluations, genetics and programming. Jan Philipsson, you are a great source of inspiration, you have shown in so many ways that science can make a difference in real life. Jette Jacobsen, thank you for all your help, especially

with finding my way through the MACE maze. Mohammad Nilforooshan thank you for your support with R, Dan Englund and Carl Wasserman, this wouldn't have been possible to do without your computer skills, thank you. Last but not least, my girls Valentina Palucci and Eva Hjerpe, thank you for all your help, support and our many laughs.

All my colleagues at the department, thank you for making this a stimulating place to work at. I have learnt a lot from many of you. Susanne Eriksson, you are really good as director of postgraduate studies, thank you for your help and support. Jörgen Sahlin, Harriet Staffans and Monica Jansson, your help with various administrative issues are greatly appreciated. The PhD-student group, past and present members, I have enjoyed spending these years with you. Lina Carlström, Emelie Zonabend König, Helena Eken Asp, Sofia Nyman and Helena Lundgren, I really found some good friends during these years, thank you so much for everything!

Mamma och pappa, stort tack för all hjälp med hund och barn! Jonas, Emilia och Filip, ni är mina bästa.