Norwegian University of Life Sciences
Faculty of Veterinary Medicine and Biosciences
Department of Animal and Aquacultural Sciences

Master Thesis 2015
30 credits

Estimation of Genomic Variance Components in Norsvin Landrace

Estimering av genomiske varianskomponenter for Norsvin Landsvin

Hoff, Solvei Cottis

Acknowledgement

My five years at the Norwegian University of Life Sciences (NMBU) are coming to an end. It has been an instructive and memorable period of my life. I grew up on a dairy farm and so the interest in animal science has been developed over many years. Animal breeding became my discipline of interest during the bachelor degree because it fascinates me how much we can change by systematic breeding, and how important it is to have a long-term perspective on the breeding work. I chose to write about genomic selection (GS) because I wanted to learn more about this "new" topic within animal breeding.

I would like to express my gratitude to my supervisors for all support during the process; Theodorus Meuwissen and Odd Vangen at the Department of Animal Science and specially Ina M. Andersen-Ranberg and Øyvind Nordbø in Norsvin. Thank you for your patience and willingness to help! I would also like to thank Ellen Rinell for proofreading my paper.

It has been a pleasure to collaborate with Norsvin. The organization has shown much interest in my work which has been very motivating.

Last but not least I would like to thank my boyfriend, family and friends for all your patience and support! Finally, special thanks to Juni S.Karlsen and Stine Samsonstuen for laughing and crying together with me this semester – we did it!

Department of Animal and Aquacutural Sciences

NMBU

Ås, 15.05.2015

…….

Solvei Cottis Hoff

Summary

Several of the major pig breeding companies have implemented genomic selection during the last years, but still there is put much effort in optimizing the use of genomic information. The Norwegian pig breeding company, Norsvin has implemented GS by using adjusted single step method (ssGBLUP), which utilize genomic relationship coefficient in addition to pedigree relationship. There has not been done any change of the variance components, which is calculated based on pedigree relationship.

The purpose of this paper is to look into consequences of implementation of GS by comparing methods with pedigree- (A), genomic- (G) and combined (H) relationship matrices for estimation of variance components and quality of the methods. There were used three different methods, best linear unbiased prediction (BLUP) which uses A-matrix as variance structure, genomic best linear unbiased prediction (GBLUP) using G-matrix and adjusted single-step (ssGBLUP) using H-matrix.

The methods were tested on four traits: growth (days from 40 to 120kg live weight), feed consumption (kg feed from 40kg to 120 kg live weight), lean meat percentage (percentage meat of carcass) and total born (still born plus live born). The first three characters were recorded at Delta, Norsvins test station for boars, for individuals born between 2011 and 2014. All individuals had both genotype and phenotype. The dataset included 4578, 4635 and 4606 individuals in the phenotype file, 6686, 6829 and 6788 in the genotype file and 12118, 12263 and 12214 in the pedigree file for growth, feed consumption and lean meat percentage, respectively. The pedigree files were seven generations deep for all four traits. Total born were recorded on sows for each farrowing in the period from January 2010 to March 2015. In total there were 129186 records registered for 62106 sows. Number of genotyped sows were only 3030, a marginal portion of the phenotyped animals. The traits were tested by univariate linear animal models.

Results from variance analysis showed no significant variation for the variance components dependent on relationship matrix. Comparison of log likelihood between methods using A- and H-matrices showed marginal better likelihood for the method using H-matrix for all traits except total born, where they were the same. The methods using genomic relationship, GBLUP and ssGBLUP, obtained similar but higher results for predictive ability than the BLUP method (e.g. correlation between predicted and observed phenotype, $corr(\hat{y}, y)$ for total born, which showed the least difference between BLUP and ssGBLUP, was 0.28 and

0.30 for BLUP and ssGBLUP respectively). Regression coefficients deviated from one for all methods, indicating that the methods were biased.

To conclude, the non-significant results for variance components indicate that there is unnecessary to estimate variance components based on genomic relationship when implementing GS. The choice of method had larger effect when estimating breeding values than estimating variance components as the GBLUP and ssGBLUP obtained better likelihood and predictive ability than BLUP. Regression coefficients showed that all methods were biased. Improved predictive ability for total born implied that adjusted single step method is a convenient way to implement GS, obtaining better predictions of breeding values for both genotyped and non-genotyped animals.

Sammendrag

Flere av de store svineavlselskapene har innført genomisk seleksjon (GS) i løpet av de siste årene, men fortsatt arbeides det mye med å optimalisere bruken. Den norske avlsorganisasjonen for gris, Norsvin, har implementert GS ved å bruke tilpasset single-step metode (ssGBLUP) som bruker genomiske slektskapskoeffisienter (G-matrise) i tillegg til tradisjonelle slektskapskoeffisienter (A-matrise). Det har ikke blitt gjort noen endring for varianskomponentene, som er beregnet på grunnlag av A-matrise.

Hensikten med denne oppgaven var å se nærmere på konsekvensene av innføring av GS ved å sammenligne metoder med tradisjonelle- (A), genomiske- (G) og kombinerte (H) slektskaps matriser for estimerte varianskomponenter og kvalitet på metodene. Det ble brukt tre forskjellige metoder, best linear unbiased prediction (BLUP) som bruker A-matrise som variansstruktur, genomisk best linear unbiased prediction (GBLUP) som bruker G-matrise og tilpasset singel-step metode (ssGBLUP) som bruker H-matrise.

Metodene ble testet på fire egenskaper: tilvekst (dager fra 40 til 120kg levende vekt), fôropptak (kg fôr fra 40kg til 120 kg levende vekt), kjøttprosent (prosentandel kjøtt av slakt) og total fødte (antall levende fødte pluss dødfødte). De tre første egenskapene ble registrert ved Delta, Norsvins teststasjon for råner, for individer som var født mellom 2011 og 2014. Alle råner hadde både genotype og fenotype. Datasettet inkluderte 4578, 4635 og 4606 individer i fenotypefilen, 6686, 6829 og 6788 i genotypefilen og 12 118, 12 263 og 12 214 i pedigreefilen for tilvekst, fôropptak og kjøttprosent, respektivt. Pedigreefilene gikk syv generasjoner tilbake for alle fire egenskapene. Total fødte ble registrert på purker for hver grising i perioden januar 2010 til mars 2015. I alt var det 129 186 registreringer fordelt på 62 106 purker. Antall genotypede purker var bare 3030, en liten andel av de fenotypede dyrene. Egenskapene ble testet enkeltvis med lineær dyremodell.

Resultatene fra varians analysene viste ingen signifikant variasjon for varianskomponentene ved bruk av ulike slektskaps matriser. Sammenligning av log likelihood mellom metodene som brukte A- og H-matriser viste marginalt bedre log likelihood for metoden som brukte H-matrise for alle egenskaper bortsett fra total fødte, hvor log likelihood var lik uavhengig av hvilken slektskapsmatrise som ble brukt. Metodene som brukte genomisk slektskap, GBLUP og ssGBLUP, fikk like, men høyere resultater for prediksjonsevne enn BLUP-metoden (f.eks korrelasjonen mellom predikert og observerte fenotype, Corr (y, y) for total fødte, som viste

den minste forskjellen mellom BLUP og ssGBLUP, var 0,28 og 0,30 for BLUP og ssGBLUP, respektivt). Regresjonskoeffisientene var forskjellig fra en for alle metoder.

Det ble konkludert med at ikke-signifikante resultater for varianskomponenter tyder på at det er unødvendig å beregne varianskomponenter basert på genomisk slektskap ved implementering av GS. Resultatene ga bedre log likelihood og prediksjonsevne for GBLUP og ssGBLUP enn for BLUP. Videre ble det derfor konkludert med at valg av metode hadde større effekt ved beregning av avlsverdier enn ved estimering av varianskomponenter. Regresjonskoeffisientene viste at alle metoder var biased. Forbedret prediksjonsevne for total fødte ved ssGBLUP indikerte at denne metoden er en praktisk metode for å implementere GS, fordi den ga gode prediksjoner av avlsverdier både for genotypede og ikke-genotypet dyr.

Table of content

1 Introduction

Applied animal genetics have played an extremely important role in the development of today's farm animals. New approaches and technologies have been utilized proportionally with the development of the different breeds. Around the 1950s farmers started to organize their breeding work in breeding associations. Within the breeding associations they systematized their work in breed specific breeding programs and constructed selection indexes (Hazel 1943). It was challenging to keep all the information from tests and pedigrees in order, but as the computer technology developed during the end of the 20th century new methods could be employed (Bourdon 2000. p.245).

For the last two decades genomic information has been the "hot topic" within genetics, promising new approaches for a variation of disciplines. Animal breeding is one of the fields of research that has had, and still has, high expectations for what can be achieved by genomics. At first, there were high expectations for Quantitative Trait Loci (QTL) experiments. By using neutral molecular markers one can locate QTLs which have influence on specific traits. Some important findings of causal polymorphisms have been made but not in significant amounts, thus it has not been directly applied but used as a complementary tool to traditional additive genetics (Ibáñez-Escriche et al. 2014). Meuwissen et al. (2001) developed new statistical methods for using genomic information to predict breeding values from markers, and in 2009 a 62k markers SNP chip called PorcineSNP60 BeadChip became commercially available for use on swine (Ramos et al. 2009). By using information from SNP chips in the newly developed methods, breeding companies were one step closer to adopt genomic selection (GS), and by now several of the major pig breeding companies have implemented GS in their breeding program (Ibáñez-Escriche et al. 2014). The aim of implementing genomic selection is to get more precise information about the animals' genetics earlier in life and in turn improve the accuracy of the EBVs/GEBVs and increase the genetic gain, especially for traits measured late in life and sex-related trait as for example maternal traits (Grindflek 2013). Improved predictions of breeding values will increase the genetic gain and so improve the economy in swine production. Although many breeding associations have implemented GS there are still much work going on to optimize the use of genomic information in the breeding work.

The Norwegian pig breeder association, Norsvin, started to look into GS in 2008. During the autumn 2011 all boars were genotyped when they had finished the phenotype test at Norsvin's boar test station, Delta, and from the autumn 2012 all boars and their mothers were genotyped when the boars were chosen as selection candidates. From January 2014 GS was implemented in their breeding work by adjusted single-step method (ssGBLUP) which is a modification of BLUP and so the only change from their traditional EBV is that they use a relationship matrix combined of genomic and pedigree relationship coefficients. Norsvin has not implemented any new method for estimation of genetic parameters and breeding values.

The aim of this paper is to look into consequences of implementation of GS by comparing methods with pedigree- (A), genomic- (G) and combined (H) relationship matrices for estimation of variance components and quality of the methods. There will be used three different methods, best linear unbiased prediction (BLUP) which uses A-matrix as variance structure, genomic best linear unbiased prediction (GBLUP) using G-matrix and adjusted single-step (ssGBLUP) using H-matrix. The methods are tested on four traits; growth, feed consumption, lean meat percentage and total born. The BLUP methods, including modifications of BLUP, like GBLUP and ssGBLUP, require genetic variance to be known to estimate breeding values. By comparing variance components estimated based on A-, H- or G matrix we can find out if type of relationship has noteworthy effect on variance components. If that is true it may indicate that it is important to calculate variance components based on genomic relationship when implementing GS. Quality of the methods will be tested by calculation of log likelihood and predictive ability.

2 Literature

2.1 Traits

2.1.1 Growth

The main purpose of swine production is to produce pork. Therefore growth traits have an important role in breeding work. Faster growth is advantageous because pigs will be ready for slaughter at earlier age, the farmer will be able to produce more slaughter pigs per year, and feed costs will be reduced. One cannot forget that growth traits also are correlated to other traits, and not all of those correlations are favorable. Referring to earlier studies (Ducos et al. 1993; van Wijk et al. 2005) growth traits are unfavorably correlated with carcass and meat quality traits.

Growth is most often measured as gram per day, but because Norsvin register growth during the test period at Delta, which endures for the time the pig takes to grow from 40 to 120kg, they measure growth as number of days from 40 to 120kg. Number of days is then the same as growth rate as the test period is over a fixed weight period. There can also be various traits of growth, for example with respect to different stages of physiological maturation. Andersen-Ranberg et al. (2013) did a study on different growth periods and found lower heritability for growth in days from 40-120 kg compared to 25-100 kg on 0.30 and 0.39 respectively. Clutter and Brascamp (1998) made a summary of results obtained from studies done between 1962 and 1998 for estimates of heritability of growth traits. The summary showed an average heritability of 0.31. Growth traits are in general considered to be moderately heritable. When estimating breeding values for Norsvin Landrace it is used a heritability of 0.39 for growth in days from 40 to 120kg (I.M. Andersen-Ranberg, personal communication, February 2015).

2.1.2 Feed efficiency

Feed is the main cost in swine production, thus feed efficiency is emphasized in the breeding work (Andersen-Ranberg et al. 2013). Favorable correlation has been found between feed conversion ratio (kg feed/kg meat) and daily gain which indicates that increased growth leads to higher feed efficiency (Ducos et al. 1993; Nguyen & McPhee 2005). Feed efficiency can be measured in several ways but most often it is measured as feed conversion ratio. In Norsvin feed efficiency is recorded as feed consumption from 40 to 120kg live weight.

In general, feed efficiency is considered as moderately heritable. In the summary table made by Clutter and Brascamp (1998), the heritability for feed efficiency traits was on average, 0.30. Gilbert et al. (2007) did a study on residual feed intake on large white and found heritability between 0.23 and 0.35 for feed efficiency traits. A heritability of 0.34 is used for feed efficiency when estimating breeding values for Norsvin Landrace (I.M. Andersen-Ranberg, personal communication, February 2015).

2.1.3 Lean meat percentage

The main income source in swine production is payment for the carcass, thus carcass traits are economically important. Carcass quality is highly valued by slaughterhouses, but also by consumers in thoughts of request of lean meat. Lean meat percentage is a measure of meat as percentage of carcass (meat, fat and bones). It is found to have moderate but unfavorable correlation to length of life (Sobczyńska et al. 2013). Another study found an unfavorable correlation between lean meat percentage and number of live-born piglets in both the first and second litter (Holm et al. 2004).

As for growth and feed efficiency there are different ways to measure lean meat percentage. The most common way is to measure lean meat percentage on slaughtered relatives of the test animals. Another way, in which Norsvin uses, is to measure lean meat percentage directly on the test animal by computed tomography (CT). The CT processes more than one thousand transverse slice images of the pig. By circulating x-rays and a detector around the body and measuring the attenuation coefficient in all pixels (small elements of the digital picture) of the image one can find the body composition of meat, bone and fat. Viscera, contents of the stomach and genitals, can be removed digitally (Kongsro 2009).

Carcass traits are known as high heritable. Ducos et al. (1993) estimated the heritability of lean meat percentage, recorded in slaughtered relatives, at 0.68 for French Landrace. Sonesson et al. (1998) also recorded lean meat percentage from slaughtered relatives and estimated heritability of 0.41. Studies of lean meat percentage estimated by computed tomography (CT-scan) has found heritability of 0.58 (Andersen-Ranberg et al. 2013) and 0.50 (Gjerlaug-Enger et al. 2012) for Norsvin Landrace. Currently, Norsvin uses a heritability of 0.55 for lean meat percentage for Norsvin Landrace (I.M. Andersen-Ranberg personal communication, February 2015).

2.1.4 Total born

Total born is considered a fertility trait and therefore it is highly weighted in the breeding goal for female lines. As for growth and feed consumption, reproduction traits also are correlated to other traits. Too large litters may lead to more stillborn piglets and decreased birth weight (Johnson et al. 1999) if the sows do not have the resources it takes to farrow and care for the piglets. The heritability of reproduction traits is known to be relatively low, between 0.08 and 0.32 (Johnson et al. 1999). Results for litter size found by Haley et al. (1988); (according to Southwood & Kennedy 1990) showed a heritability of 0.10 and in the summary table made by Rothschild and Bidanel (1998) heritability for total born is 0.11. The heritabilities used for Norsvin Landrace is 0.09 for total born (I.M. Andersen-Ranberg personal communication, February 2015).

2.2 Registration and use of the traits in Norsvins breeding work

The traditional way to select breeding candidates is to select based on phenotypes for traits of interest. Animals perform differently for same trait because of variation in genotypes and/or environment, thus it is possible to select the best ones. The main purpose of breeding programs is to obtain genetic gain and so improve the economy for current production. Hence traits in Norsvins breeding goal are of high economic value in swine production. However, to obtain genetic gain is challenging because many factors play a role, for example selection intensity, accuracy, degree of inbreeding, genetic standard deviation and generation interval (Bourdon 2000).

Norsvin has organized their breeding work in a "breeding pyramid" with two main levels; nuclei and multiplier herds. In addition they have a test station, Delta, where they test all AI boar candidates (Aasmundstad et al. 2014a). The breeding nuclei produce the next generation of elite animals and supply the test stations with boars and the multipliers with sows. Selection of the breeding animals is organized differently for gilts and boars. For gilts, selection is done by the farmer himself based on information from a farm test, registrations in the data system "InGris" and the farmer's general impression of the gilt. Trained technicians perform the farm test. The test involves measure of lard and muscle thickness by ultrasound, registration of weight and number of inverted teats and an evaluation of conformation and mobility. All gilts born in the nucleus- and multiplier herds, circa twenty thousand per year, are tested at around 150 days old. Farmers register information on reproduction (total born

e.g.), health and appearance in “InGris”. “InGris” is the national management and registration tool for pig production. It is web based, and all data are collected in a central database (Aasmundstad et al. 2014b). Around 2800 Norsvin Landrace sows are used for breeding annually. The actual boar selection candidates are selected based on EBV made on litters and evaluation of individual exterior.

Each year around 3500 boars are tested at Delta, of which approximately 1200 are Norsvin Landrace. Roughly 60 out of the 1200 boars (top 5%) are selected to be elite boars. Selection candidates arrive at the test station when they are about 40 kg. The phenotype testing starts once they arrive and lasts until they reach 120kg. The boars are stalled in pens of 12 individuals. In each pen there is a FIRE-station which is an automatic feeding station and scale. The pigs have ad libitum feeding and each time a pig enters the FIRE-station feed consumption, time of each visit, number of visits and body weight are recorded. When the boars have reached 120 kg and finished the test, they are further tested using CT. The CT gives a lot of information which, among other things, is used to estimate bone quality and carcass traits. After the results from the phenotype test and the CT-scan are ready the boar gets a selection value.

The breeding goal for Norsvin Landrace is presented in figure 1 (Norsvin 2013). There are 26 traits, divided into seven different categories that contribute to the breeding goal. Norsvin uses Norsvin Landrace as a female line and therefore robustness, litter size and maternal ability have the heaviest weightings.

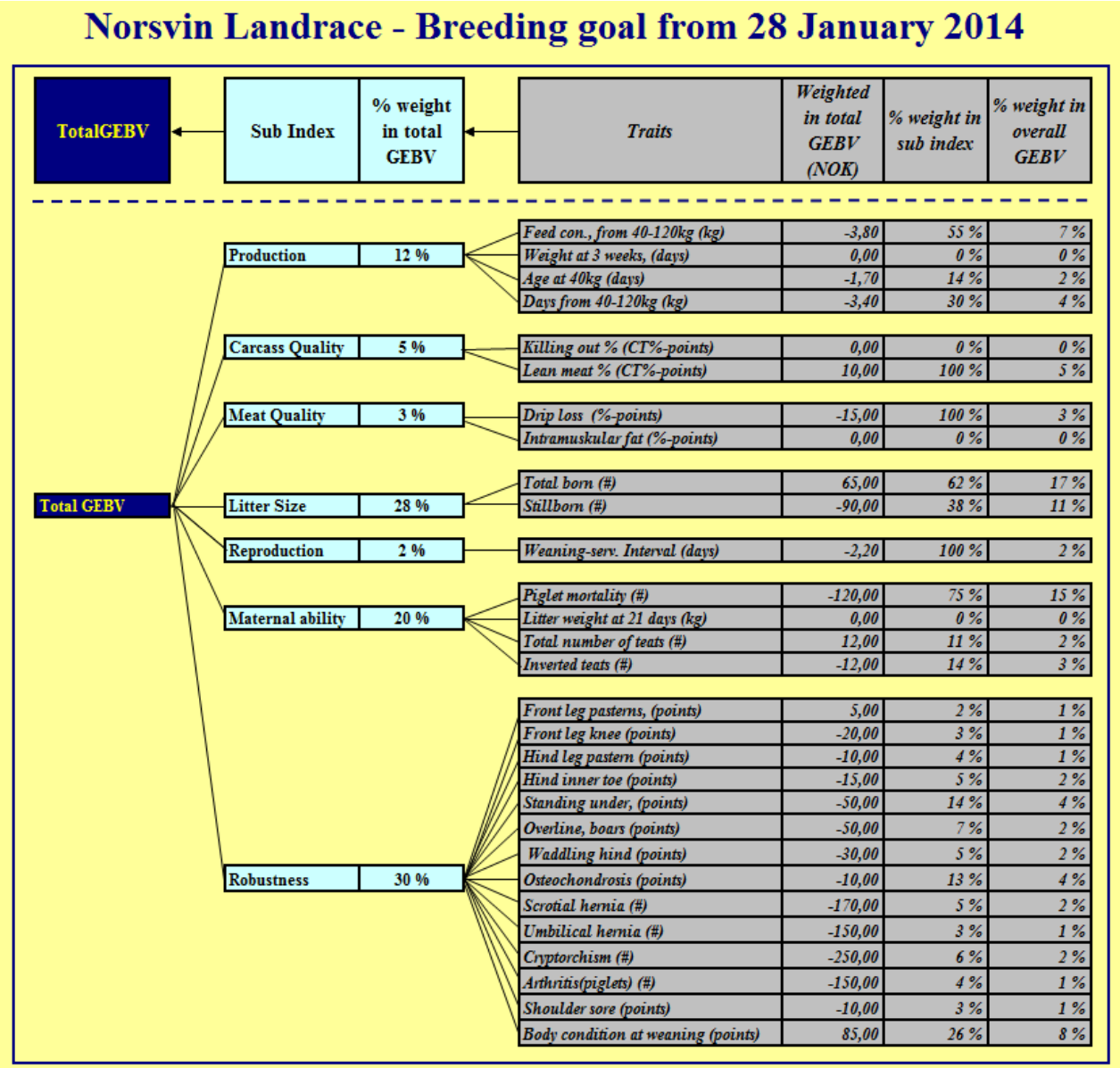


Figure 1: Breeding goal for Norsvin Landrace from 28.01.2014 with all traits and their weightings (Norsvin 2013).

The breeding goal shows the weighting of the different traits. The production category consists of four traits which have a total weighting of 12% in the breeding goal. The four traits are: feed consumption, weight at 3 weeks, age at 40 kg and days from 40-120 kg with 7%, 0%, 2% and 4% weight for each trait respectively. Carcass quality is weighted at 5% with all weight on lean meat percentage. For the category litter size there are two traits in the breeding goal; total born weighted at 17% and stillborn at 11%, making a total of 28% for litter size of the total breeding goal.

2.3 Genomic selection

Best linear unbiased prediction, BLUP, is the traditional method to estimate breeding values based on phenotypes and traditional relationship. During the last decade many different methods have been developed in attempt to optimize the implementation of genomic information for estimation of genomic estimated breeding values (GEBVs). By using DNA-chip technology, the effect of thousands of DNA markers can be analyzed simultaneously. Single nucleotide polymorphisms (SNPs) are currently the most common genetic markers. Meuwissen et al. (2013) describes genetic markers as:"Loci whose alleles can be used to keep track of a chromosome or a chromosomal region during the transmission from parent to offspring". If a single nucleotide at the same position on the genome is different between two animals then it can be used as a SNP marker. SNP-chips, comprised of thousands of SNPs distributed throughout the whole genome have shown to be a cost-effective method to genotype animals, and it increases genetic gain compared to conventional breeding methods (Meuwissen et al. 2001).

Meuwissen et al. (2001) put forward one method to implement GS in breeding work; genome-wide selection, also called multi-step approach. This method requires reference populations with animals having both phenotype and genotype records. GEBVs are estimated by comparing one individual's genotype with the estimated marker effects in the reference population. Another method is the single-step approach which uses modifications of the traditional BLUP method, changing the pedigree relationship matrix by relationship matrices including genomic relationship coefficients (Christensen et al. 2012; Legarra et al. 2009). The single-step method is easier to handle and require less computer technology than multi-step method (Gao et al. 2012).

Breeding organizations mainly utilize one of the two above mentioned methods to implement GS. In this study the single-step method will be used since that is the method which Norsvin make use of. Thus the only change from BLUP is that the variance structure is based on a combination of genomic and pedigree relationship matrices and not only pedigree relationship as earlier. The traditional pedigree relationship is based on statistical likelihood; e.g. full-sibs are 0.5 related to each other and half-sibs 0.25. Genetic relationship is more precise and in practice it has been shown that the relationship between half sibs varies from 15-35% (Lopes et al. 2013). The implementation of GS by single-step method is therefore expected to lead to

increase in genetic progress, especially for the female lines in which most trait records are done late in life and cannot be recorded on boars (Ibáñez-Escriche et al. 2014).

There are several modifications of the single-step method. One method is genomic best linear unbiased prediction, GBLUP. In this method the pedigree-based relationship matrix (A) in the animal-based model is replaced by a genomic relationship matrix (G). GBLUP is not suitable for predictions of non-genotyped animals since it only utilizes genomic relationships in the calculations of GEBVs. Another method is the original single-step method, which uses a combination of pedigree and genomic relationships by making a combined relationship matrix (H). By combining G- and A-matrix information from both genotyped and non-genotyped animals can be used and it is possible to predict GEBVs for all animals independent of genotype. Because values in the A- and G-matrix often can be on different scale, there is developed an advanced version of the single-step method called "adjusted single step method" in which the G-matrix values are adjusted to the A-matrix values (Christensen et al. 2012). Christensen et al. (2012) compared BLUP, GBLUP and original and adjusted single-step methods for genetic gain and feed conversion ratio in Danish Duroc pigs. Overall the results showed highest accuracy for adjusted single-step methods also when it came to dataset with many non-genotyped animals. The GBLUP method for genotyped animals and the original single-step method (without adjusting of G- to A-matrix values) were more accurate than BLUP. Norsvin utilize adjusted single-step method in their breeding work. In the current study it is abbreviated ssGBLUP.

3 Materials and methods

3.1 Material

3.1.1 Animals and phenotypes

Records used in this study are retrieved from the routine run for EBVs in the breeding population of Norsvin Landrace. Thus, Norsvin has tested records and trait models so there was no need to test the models or check the dataset for outliers etc. Number of individuals in the phenotype-, genotype- and pedigree files is presented in table 1. The files were processed in SAS software (SAS 2013). All pedigree files were constructed using a seven generations deep pedigree.

There were registered phenotypes for four traits; growth (days from 40 to 120kg live weight) feed consumption (kg feed from 40kg to 120 kg live weight), lean meat percentage (percentage meat of carcass) and total born (still born plus live born). Growth, feed consumption and lean meat percentage were measured on boars during the phenotype test at Delta. From autumn 2011 all tested boars have been genotyped, thus the animals have both phenotype and genotype. Data included boars born between 2011 and 2014. There were only used boars with both phenotypes and genotype in the present study, hence non-genotyped boars from 2011 were not used. Boars born in 2011 and 2012 were merged into one year group to ensure a proper size for the year group. This resulted in three year groups in the datasets for traits recorded at Delta. Number of phenotyped- and genotyped individuals was around 4600 and 6700, respectively (table 1). Norsvin did genotype some boars before 2011 in condition with research, thus total number of genotyped animals was higher than number of animals with recorded phenotypes. Growth and feed consumption were registered in the FIRE-station and lean meat percentage was obtained by CT-scan at Delta.

Total born was registered in “InGris” for the sows in the nucleus herds. The phenotypes were retrieved from the routine run for EBV in the period from January 2010 to March 2015. Norsvin changed their parameter of litter size, in 2010, from live born to total born and still born, thus the phenotypic measures have been equal and of proper quality for the records in the dataset for total born. The phenotype file contained 62106 individuals. Only sows having a son that has been sent to Delta are genotyped, thus only a sparse amount of the animals are genotyped (3030 individuals in the genotype file).

Table 1: Number of individuals for the phenotype, genotype and pedigree files for growth, feed consumption, lean meat percentage and total born.

Traits	Phenotype file	Genotype file	Pedigree file
Growth	4578	6686	12118
Feed consumption	4635	6829	12263
Lean meat percentage	4606	6788	12214
Total born	62106	3030	72392

Table 2 presents number of observations, average value, standard deviation and minimum- and maximum values for the four traits. Number of observations was around 4600 for the traits tested at Delta (growth, feed consumption and lean meat percentage), while it was 129186 for total born. Total born is a repeatability model which mean that one individual can be registered for same trait several times, therefore number of observations (129186, table 2) is many more than number of individuals in the phenotype file (62106, table 1) The average values obtained from the actual datasets were 75.87days for growth from 40 to 120kg live weight, 173.14kg feed consumed during the test period (40 to 120kg live weight), 65.21% lean meat of carcass and 13,81 total born per sow.

Table 2: Number of observations, average value, standard deviation, minimum and maximum value for growth, feed consumption, lean meat percentage and total born used in the analysis.

	N-obs	Average	S.D	Min value	Max value
Growth	4578	75.87	6.98	58.49	119.32
Feed Consumption	4635	173.14	12.38	134.72	236.23
Lean meat percentage	4606	65.21	3.35	51.98	76.56
Total born	129186	13.81	3.57	1.00	30.00

3.1.2 Genotypes

In Norsvin's breeding system all boars at Delta and all dams who have a son that was sent to the test station are now genotyped (from January 2015 all pregnant sows in nuclei herds are genotyped). A tissue sample from their ear is sent to the Biobank (Hamar, Norway) where they isolate the DNA and send the sample to CIGENE (Aas, Norway)(Grindflek 2013). As described in Aasmundstad (2014) the DNA are genotyped: "using the iScan (Illumina, San Diego, CA, USA) platform with the PorcineSNP60 array according to the manufacturer's instructions. Image intensity data processing, clustering and genotype calling were performed using the genotyping module in the Genome Studio software" (Illumina, San Diego, CA, USA). "All SNP markers have to pass a quality control before the genotyped animals are included in the genomic relationship matrix." The control parameters are; minor allele frequency > 0.01, call frequency > 95% and Parent-Child Mendelian errors < 1% (H. Hamland, personal communication, May 2015). In current study the registered genotypes for animals of interest were retrieved from Norsvins database.

3.2 Methods

3.2.1 Estimation of variance components

The genomic relationship matrices and the inversed genomic relationship matrices were made in DMU built by using GMATRIX (Su & Madsen s.a.). The program required a marker file, a map file and a file with IDs to the animals of interest. The marker file contained the genotype for the animals. Norsvin processes large marker files for all genotyped animals, thus by merging the ID file containing ID's of animals of interest with the large map file the model specific marker file was made. The map file contained information about locus and was used to calculate the number of marker loci. Since the map file is breed specific the same file was used for all analyses. The options for the GMATRIX program are contained in the parameter file. The same options as Norsvin uses were used as follows; minor allele frequency of 0.01, marker allele frequency was calculated from data, the G-matrix was scaled by M-matrix divided by sqrt(2pq) for each locus and the G-matrix values were not adjusted to the same scale as for the A-matrix (adjusting was done in DMU when it was necessary).

The variance analysis was conducted by using AI-REML algorithm using DMUAI in DMU, 6 software package, release 5.2 (Madsen & Jensen 2013). Norsvin uses multi-trait model, but in

the current study each trait was analyzed by univariate animal models to simplify the computations and handling of data, consequently covariance between traits are not taken into consideration. The option under AI-REML was chosen based on what kind of relationship that was used. The PED option in DMU was used for the analysis based on traditional relationship (A matrix). For the analysis with pedigrees combined by traditional and genomic relationship (H-matrix) the PGMIX option in DMU was used. H-matrix was made with 0% weighting of A-matrix and 100% of G-matrix, meaning pedigree relationship coefficients were only used if the genomic relationship coefficients were missing. To test if different weightings of A-matrix into H-matrix would have improved the method, there were performed variance analysis with weighting of A-matrix with 10, 20, 30, 40 and 50% to obtain the likelihood. Different weighting was only tested for the growth trait. Elements of the G-matrix were adjusted to the corresponding pedigree relationship matrix. For the genome-based relationship, the GREL option in DMU was used. Because of very few genotyped animals for total born, there were not made solutions of GREL for that trait. Log likelihood was generated for all the variance analysis. As mentioned earlier, the trait models are the same as used by Norsvin. They are mixed models.

Mixed model in general form;

$$y = X\beta + Z\mu + e \quad (1)$$

Where;

y = the observed phenotype for each animal.

β = a vector of fixed effects with design matrix X.

μ = a vector of random effects with design matrix Z.

e = error term

The models (in detail) for each trait:

$$Growth_{ijklmnop} = H_i + B_j + PM_k + S_l + PE_m + PA_n + I_o + LB_p + e_{ijklmnop} \quad (2)$$

Where;

Fixed effects:

H= herd and year (i=1,….88)

B= month of birth (j= 1,….12)

PM= parity mother (k= 0, 1, 2, 3)

S= section at Delta (l=1,…83)

Random effects:

PE= pen at Delta

PA= parity

I= animal

e= error term

Fixed regression effects:

LB= number of live born piglets in litter (p=1,….23)

$$Feed\ consumption_{ijklmnop} = H_i + B_j + PM_k + S_l + PE_m + PA_n + I_o + LB_p + e_{ijklmnop} \quad (3)$$

Where $H_i, B_j, PM_k, LB_p\ and\ e_{ijklmnop}$were the same as for growth, while the rest of the variables and levels were:

Fixed effects:

S= section at Delta (l=1,…86)

$$Lean\ meat\ percentage_{ijklmnoqp} = H_i + B_j + PM_k + S_l + PE_m + PA_n + I_o + LW_q + LB_p + e_{ijklmnoqp} \quad (4)$$

Where $H_i, B_j, PM_k, LB_p\ and\ e_{ijklmnop}$were the same as for growth, while the rest of the variables and levels were:

Fixed effects:

S= section at Delta (l=1,….81)

Fixed regression effects:

LW= body weight at CT-scan (q=1,…2167)

Random effects:

$e_{ijklmnoqp}$= error term

$$Total\ born_{kvstuwhz} = MP_k + PN_v + HY_s + SF_t + BY_u + MPN_r + I_0 + IC_w + AM(MPN)_h + AM2(MPN)_z + e_{kvsturowhz} \quad (5)$$

Where;

Fixed effects were:

MP= mothers parity (k= 0, 1, 2, 3)

PN= parity (v=1,…7)

HY= herd and year of birth (s=1,…..552)

SF= season of farrowing (t=1,…4)

BY= breed and year of birth (u=1,….63)

Random effects were:

MPN= mothers ID and mothers litter number

I= animal

IC= genetic effect with a covariance matrix proportional to the relationship matrix.

$e_{kvsturovwz}$= error term

Fixed regressions effects:

AM(MPN)= mothers age at farrowing (h=1,….1003)

AM2(MPN)= mothers age at farrowing^2 (z=1,….1003)

AM_h and $AM2_z$ were nested with MPN_r. When processing levels for MP (mothers parity) parities over 3 were set as 3. This was not done for PN (parity).

Heritability was calculated as the ratio of individual variance divided by the sum of variance for the random effects. The formula of heritability for growth, feed consumption and lean meat percentage was:

$$h^2 = (var(I)/(var(I) + var(PA) + var(PE) + var(e)) \quad (6)$$

The formula of heritability for total born;

$$h^2 = ((var(I)/(var(I) + var(IC) + var(MPN) + var(e)) \quad (7)$$

3.2.2 Testing predictive ability

All data analyses for predictive ability were analyzed using DMU. To test the predictive ability, three different parameters were calculated: correlation between predicted and observed phenotype, $corr(\hat{y}, y)$, correlation between EBV and observed phenotype $corr(\hat{\mu}, y)$ and correlation between EBV and observed phenotype with correction for environmental effects, $corr(\hat{\mu}, y - X\hat{\beta})$ (here $X\hat{\beta}$ include solutions for non-genetic random effects in addition to fixed effects) . The datasets for each trait were divided into a test set and a validation set. The size of the validation set was around 400 individuals (randomly picked out from the 800 youngest individuals in the data set) because it corresponds approximately to a standard error of a correlation estimate of 0.05. (The standard error of a correlation estimate is se(r)=sqrt((1-r^2)/(n-2)) is approx. sqrt(1/n). Where r= correlation and n= number of individuals). An animal was included in the validation only if there were solutions for all fixed effects registered on the animal, hence the size of the validation sets varied from 391 to 551 records. The validation set with 551 records is for the trait total born, which is a repeatability model, thus one individual can have plural records. EBVs or GEBVs were estimated for all animals in the test set. For animals in the validation set, EBVs and GEBVs were predicted by using phenotypes, relationship and variance components from the test set. Afterwards, the predicted phenotypes were compared to the masked phenotypes. Five models were tested for each trait; BLUP, GBLUPg, GBLUPa, ssGBLUPg and ssGLUPa. The models differed in base for relationship matrix, variance structure and if the variance components were estimated with A-, H- or G- relationship matrix (see table 3). BLUP utilize pedigree relationship as covariance structure and the variance components are based on A-matrix. GBLUP employ genomic relationship matrix as covariance structure. The little -g and –a indicate if the variance components were estimated based on G-matrix or A-matrix, respectively. This applies to the ssGBLUP method too. ssGBLUP utilize combined

relationship matrix as covariance structure. The breeding values were then estimated by DMU5 for BLUP, ssGBLUPg and ssGBLUPa and by DMU4 for GBLUPg and GBLUPa (see table 3).

Table 3: Overview of the different methods (BLUP, GBLUPg, GBLUPa, ssGBLUPg and ssGBLUPs) tested for predictive ability.

	BLUP	**GBLUPg**	**GBLUPa**	**ssGBLUPg**	**ssGBLUPa**
Relationship matrix	A-matrix	G-matrix	G-matrix	H-matrix	H-matrix
Basis for variance components	A-matrix	G-matrix	A-matrix	H-matrix	A-matrix
DMU module for EBV	DMU5	DMU4	DMU4	DMU5	DMU5

Once EBVs were obtained for all models the solution files were divided into individual files for fix-, random- and regression effects and one file containing all animals with their registered phenotypes and EBV/GEBV. In cases where any solutions for fixed effects registered on the animal were not obtained, the animal was excluded from the validation set. If any solution of random effects on the animals was not obtained, the relevant random effect was set equal to zero. The files were used in the calculation of parameters presenting predictive abilities and regression coefficients.

4 Results

4.1 Variance analysis

Estimated variance components based on A-, H- and G-matrix for growth, feed consumption, lean meat percentage and total born are shown in table 4, 5, 6 and 7, respectively. The results for each method varied, but there were no significant differences for any of the traits. When comparing the variance components obtained for the three methods, the difference tended to be largest between the method with pedigree relationship and the two methods including genomic relationship. The difference between the model using combined relationship (H-matrix) and genomic relationship (G-matrix) was very small and inconsistent. There seems a trend for all traits that residual variance increases when comparing the single-step method to the BLUP method.

The results for growth in table 4 show a trend of decreased heritability when genomic relationship is included. The standard errors of variance are reduced for individual variance when using genomic relationship, while they are unchanged for pen variance and increased for litter- and residual variance.

Table 4: Results from variance analysis for growth. The results show estimated variance for random effects and heritability for growth using two different methods for relationship matrix. Values in parentheses are standard errors of variance.

	Growth		
Relationship matrix	**A**	**H**	**G**
Individual (I)	14.576 (2.124)	12.250 (1.160)	11.813 (1.115)
Litter (PA)	11.957 (5.787)	7.073 (6.127)	6.990 (6.137)
Pen (PE)	3.819 (0.590)	4.170 (0.590)	4.169 (0.590)
Residual (e)	16.124 (5.849)	22.675 (6.122)	22.729 (6.132)
Heritability	0.314	0.265	0.258

When analyzing the trait feed consumption the method using A-matrix gave highest heritability and standard errors of the variances (table 5).

Table 5: Results from variance analysis for feed consumption. The results show estimated variance for random effects, and heritability for feed consumption using three different methods for relationship matrix. Values in parentheses are standard errors of variance.

	Feed consumption		
Relationship matrix	**A**	**H**	**G**
Individual (I)	42.606 (6.060)	36.960 (3.505)	35.364 (3.349)
Litter (PA)	27.326 (17.365)	27.349 (15.725)	27.118 (15.780)
Pen (PE)	21.907 (2.524)	22.715 (2.519)	22.699 (2.518)
Residual (e)	50.067 (17.520)	56.013 (15.659)	56.205 (15.715)
Heritability	0.300	0.258	0.250

Table 6 showed no change in heritability for the different methods for lean meat percentage. Standard error of genetic variance was highest when using pedigree relationship alone.

Table 6: Results from variance analysis for lean meat percentage. The results show estimated variance for random effects and heritability for lean meat percentage using three different methods for relationship matrix. Values in parentheses are standard errors of variance.

	Lean meat percentage		
Relationship matrix	**A**	**H**	**G**
Individual (I)	2.858 (0.343)	3.033 (0.214)	2.913 (0.205)
Litter (PA)	0.935 (0.681)	0.568 (0.559)	0.572 (0.560)
Residual (PE)	1.975 (0.698)	2.409 (0.560)	2.401 (0.560)
Heritability	0.495	0.505	0.495

The results of variance analysis for total born showed increased variance for the method using H-matrix (see table 7). Standard errors and heritability changed minimally.

Table 7: Results from variance analysis for total born. The Results show estimated variance for random effects and heritability for total born using three different methods for relationship matrix. Values in parentheses are standard errors of variance.

	Total born	
Relationship matrix	**A**	**H**
Individual (I)	1.060 (0.061)	1.072 (0.062)
Individual (IC)	0.849 (0.056)	0.860 (0.056)
Litter (PA)	0.129 (0.030)	0.141 (0.031)
Residual (e)	9.845 (0.050)	9.847 (0.050)
Heritability	0.089	0.090

4.2 Log Likelihood and Weighting of A-matrix:

Log likelihood for the methods used to estimate variance components are listed in table 8. The values for total born are closest to zero and thus represent the highest likelihood. High likelihood denotes high quality of the method. The likelihood was higher when using the method with combined relationship (H-matrix) for all traits except total born for which there was no difference.

Table 8: **Log likelihood (-2logl) for the methods using pedigree and combined relationship for the different traits.**

Relationship Matrix	**A**	**H**
Growth	-2.146*	-2.116*
Feed Consumption	-2.647*	-2.619*
Lean Meat Percentage	-1.199*	-1.136*
Total born	-4.444**	-4.444**

*E+4+constant

**E+5+constant

Log likelihood for methods weighting A-matrix from 10 – 50 % in H-matrix showed very little change. The test was only done for growth. Weighting of A-matrix by 0-30% gave log likelihood of -0.2116E+04+constant, 40% gave log likelihood of -2.117 E+04+constant and

50% gave log likelihood of -2.118 E+04+constant. The results showed little difference in likelihood, but it seemed like the likelihood decreased the more the A-matrix is weighted.

4.3 Predictive ability

The parameters representing predictive ability for the different methods and traits are shown in table 10. Predictive ability was noticeably lower for the BLUP method compared to the other methods while there were no difference between the GBLUP and ssGBLUP methods or between the methods using only genomic relationship, GBLUPa and GBLUPg, compared to those using a combination of A- and G relationship matrices, ssGBLUPa and ssGBLUPg. Considering the two different measures of predictive ability, correlation between EBV/GEBV and observed phenotype adjusted for environmental effects, $corr(\hat{\mu}, y - \hat{\beta}X)$ gave the lowest predictive ability.

Regression coefficients are presented in table 10. All regression coefficients deviated from one. Regression coefficients for observed phenotype on predicted phenotype, $b(y\ on\ \hat{y})$, showed quite similar results for all methods, but there were higher values for GBLUPg and ssGBLUPg for growth and feed consumption. Regression coefficients for observed phenotype adjusted for environment on estimated breeding values, $b((y - \hat{\beta}X)on\ \hat{\mu})$, showed only small differences between the various methods including genomic selection, but the values for the BLUP method were lower than the other methods for growth and lean meat percentage and higher for feed consumption and total born.

Table 9: Predictive ability and regression coefficients for different methods for estimation of breeding values tested on the traits: growth, feed consumption, lean meat percentage and total born. Predictive ability is presented in two parameters; correlation between predicted and observed phenotype, $corr(\hat{y}, y)$) and correlation between EBV and observed phenotype adjusted for environmental effects, $corr(\hat{\mu}, y - \hat{\beta}X)$. Regression coefficients for the same parameters are presented as $b(y\ on\ \hat{y})$ and $b((y - \hat{\beta}X) on\ \hat{\mu})$. Differences between the methods are explained in table 3.

Growth	**BLUP**	**GBLUPa**	**GBLUPg**	**ssGBLUPa**	**ssGBLUPg**
$corr(\hat{y}, y)$	0.28	0.36	0.36	0.36	0.36
$corr(\hat{\mu}, y - \hat{\beta}X)$	0.06	0.22	0.22	0.22	0.22
$b(y\ on\ \hat{y})$	0.57	0.57	0.59	0.57	0.59
$b((y - \hat{\beta}X) on\ \hat{\mu})$	0.27	0.53	0.57	0.54	0.57
Feed Consumption					
$corr(\hat{y}, y)$	0.39	0.44	0.44	0.44	0.44
$corr(\hat{\mu}, y - \hat{\beta}X)$	0.20	0.27	0.27	0.27	0.27
$b(y\ on\ \hat{y})$	0.72	0.72	0.73	0.72	0.73
$b((y - \hat{\beta}X) on\ \hat{\mu})$	0.90	0.69	0.73	0.69	0.72
Lean meat percentage					
$corr(\hat{y}, y)$	0.45	0.61	0.61	0.61	0.61
$corr(\hat{\mu}, y - \hat{\beta}X)$	0.27	0.50	0.50	0.50	0.50
$b(y\ on\ \hat{y})$	0.60	0.77	0.77	0.77	0.77
$b((y - \hat{\beta}X) on\ \hat{\mu})$	0.96	0.97	0.97	0.98	0.97
Total born					
$corr(\hat{y}, y)$	0.28	-	-	0.30	0.30
$corr(\hat{\mu}, y - \hat{\beta}X)$	0.16	-	-	0.19	0.19
$b(y\ on\ \hat{y})$	1.13			1.06	1.06
$b((y - \hat{\beta}X) on\ \hat{\mu})$	1.18			1.03	1.03

Appendix 1 presents plots of the correlation between predicted and observed phenotype, $corr(\hat{y}, y)$. There is one plot demonstrating the correlation when using BLUP and one when using ssGBLUPg for each trait. The other methods had very similar results to ssGBLUPg, thus they were not presented in this paper.

The figures in appendix 2 show the distribution of random-, fixed- and regression effects and EBV/GEBV estimated based on relatives for animals in the validation set. The animals' predicted phenotypes are based on the sum of these values. It seems like the ssGBLUPg method, which includes genomic information show more variance in GEBVs than the BLUP method do for EBVs for all traits. On the other hand, distribution of random, fixed and regression effects appeared to be alike for the two methods for all traits except for lean meat percentage, where fixed effects and regression effects seems to have less effect when using ssGBLUP. Predicted phenotypes for growth are, from looking at the first two figures in appendix 2, mainly explained by the fixed effect while the other effects range from negative five to three. Fixed effects explain the most for feed consumption and lean meat percentage too, but the regression effect negatively impacted the phenotypic expression for lean meat percentage. Regression effects explained more than the fixed effects in phenotypes for total born.

Table 11, 12, 13 and 14 show correlations of estimated EBVs/GEBVs for the animals in the validation set between the different methods for growth, feed consumption, lean meat percentage and total born, respectively. The results showed high correlations between the methods including genomic relationship, which means that they predict approximately the same value for EBVs/GEBVs. The correlation between BLUP and the other methods were lower. Standard deviation was low when using BLUP compared to the other methods.

Table 10: Standard deviations of estimated breeding values are presented on diagonal and correlations between the different methods, tested on estimated breeding values for growth, are presented of diagonal.

	Growth				
Method	**BLUP**	**GRELa**	**GRELg**	**ssGBLUPg**	**ssGBLUPa**
BLUP	*1.406*				
GRELa	0.486	*2.607*			
GRELg	0.501	0.998	*2.422*		
ssGBLUPg	0.500	0.998	1.000	*2.426*	
ssGBLUPa	0.488	1.000	0.999	0.999	*2.585*

Table 11: Standard deviations of estimated breeding values are presented on diagonal and correlation between the different methods, tested on estimated breeding values for feed consumption are presented of diagonal.

	Feed consumption				
Method	**BLUP**	**GRELa**	**GRELg**	**ssGBLUPg**	**ssGBLUPa**
BLUP	*2.458*				
GRELa	0.550	*4.246*			
GRELg	0.558	0.998	*3.961*		
ssGBLUPg	0.558	0.998	1.000	*3.965*	
ssGBLUPa	0.551	1.000	0.999	0.999	*4.199*

Table 12: Standard deviations of estimated breeding values are presented on diagonal and correlation between the different methods, tested on estimated breeding values for lean meat percentage are presented of diagonal

	Lean meat percentage				
Method	**BLUP**	**GRELa**	**GRELg**	**ssGBLUPg**	**ssGBLUPa**
BLUP	*0.725*				
GRELa	0.587	*1.284*			
GRELg	0.587	1.000	*1.283*		
ssGBLUPg	0.587	1.000	1.000	*1.283*	
ssGBLUPa	0.588	1.000	1.000	1.000	*1.273*

Table 13: Standard deviations of estimated breeding values are presented on diagonal and correlation between the different methods, tested on estimated breeding values for total born are presented of diagonal.

	Total born		
Method	**BLUP**	**ssGBLUPg**	**ssGBLUPa**
BLUP	*0.429*		
ssGBLUPg	0.742	*0.583*	
ssGBLUPa	0.743	1.000	*0.582*

5 Discussion

5.1 Variance and heritability

The variance analyses showed no significant difference in variance of random effects estimated with the different relationship matrices, A, H and G (table 4, 5, 6 and 7). However there was a tendency that the variance components changed when implementing GS, resulting in a trend of decreased heritabilities when using H- and G relationship matrices instead of A-matrix.

As the results for variance analysis were non-significant there is dubiously that it is necessary to estimate variance components based on genomic relationship when implementing GS. The accuracy would possibly have been better if multi-trait models were used instead of single-trait models. Single-trait models were used for the current study because they have lower computer requirements, and are easier to work with. However, multi-trait models are more accurate because they analyses all data at the same time. Therefore more information is included in the analysis, especially information about covariance between traits is provided. If one trait has few records, information from other correlated traits can increase the accuracy. The datasets and number of records should also be considered when thinking of accuracy of predictions. For example, number of records per litter is very low in the present study because most often only one boar is selected from each litter. This leads to less accurate predictions of the effect of litter, especially when using single trait models with no information from correlated traits. There have been performed studies testing both single- and multi-trait model, in which adjusted single step method resulted in more accurate predictions for both models (Christensen et al. 2012). For the current study there is still questionable if a multi-trait model would have given significant results for variance components.

From table 4, 5, 6 and 7 it seems like a trend that estimated heritabilites decreased for variance components based on H- and G-matrix compared to A-matrix (e.g. for feed consumption heritability changed from 0.30 to 0.26 and 0.25 for A-, H and G-matrix respectively). What is noteworthy is that Aasmundstad (2014) found increased heritability when using H-matrix compared to A-matrix. Since we expect more precise prediction with genomic relationship it is easy to assume that the heritability also should increase, but this is not necessarily true (e.g. the sire model is less accurate than the animal model, but the heritability estimates are very similar). A lower heritability for the G-matrix may be expected if one considers that both G- and A-matrix estimate the genetic variance in the founder

population, but for G the founder population is the genotyped population whereas for A the founder population lived at the beginning of pedigree recording (Veerkamp et al. 2011). Thus, the loss of genetic variance that occurred since the beginning of pedigree recording can explain the reduction in heritability when moving from A to G matrix based estimates.

The estimated heritabilities in the present study are similar to what has been reported in literature, but slightly lower, for all traits except total born, compared to the heritabilities used by Norsvin for the EBV for Norsvin Landrace. However, there was only used a small portion of all information included in the EBV by Norsvin in the current study. Because of smaller datasets and single trait models in the present study the heritability estimates may differ.

Even though there were no significant differences for estimated variance for random effects, there was a tendency that the difference was largest between the method with pedigree relationship and the two methods including genomic relationship. This is consistent with other studies (Aasmundstad 2014; Forni et al. 2011; Veerkamp et al. 2011). The reason why the largest difference tend to be between the BLUP method and the two methods including genomic relationship is most likely because the methods uses different relationship coefficients. Genomic relationship coefficients are more precise than pedigreed relationship coefficients because genetic information gives better accuracy of Mendelian sampling terms (Aasmundstad 2014; Forni et al. 2011; Veerkamp et al. 2011). The effect of implementing genomic information regarding standard errors of the variances of random effects varied for the different traits. For growth, standard error of individual variance decreased when including genomic relationship, while the standard errors of pen were unchanged. Standard errors of litter and residual variance were lowest when using BLUP method. For total born there was no change in standard errors. Feed consumption and lean meat percentage had the highest standard errors when using the BLUP method. Aasmundstad (2014) and Forni et al. (2011) also got higher standard errors when using pedigree relationship compared to a combination of genomic and pedigree relationship. Since genomic relationship coefficients are more precise it might be that the variation actually has larger variance, in which will increase the standard error.

For total born, estimated variance of the random effects varied little among BLUP and ssGBLUP. This is probably because only a small portion of the phenotyped animals were genotyped, thus little information is added when including genomic relationship coefficients. Implementation of GS is expected to lead to increased genetic gain, especially for maternal

traits which are measured late in life and on sows, not directly on the test boars. By including genotypes from sows a larger amount of the phenotyped animals will be genotyped and the accuracy of selection will increase even more (Lillehammer et al. 2011). Norsvin has taken this into consideration so from January 2015 they genotype all pregnant sows in the nuclei herds. In this way the GEBVs also can constitute a part of the selection of sows.

There were no differences in estimated variance of random effects for the methods using H-matrix compared to G-matrix. This was most likely because the traits tested were recorded at Delta, consequently all individuals were both phenotyped and genotyped. The same phenotypic information will be included in both methods when all phenotyped animals are genotyped. There could have been some difference between GBLUP and ssGBLUP regarding that the G-matrix values were adjusted to A-matrix in PGMIX option while there was no adjustment of G- to A-matrix values in the GREL option of DMU, but this seems not to be the case in the current study.

For all traits, residual variance tended to increase when comparing the method using H-matrix with the one using A-matrix. This may indicate that the estimates based on pedigree relationship are quite good while the methods including genomic relationship explain less of the variance (the SNPs might not catch all variance). On the other hand, genomic relationship coefficients, as already mentioned, are said to be more precise.

5.2 Log likelihood

Log likelihood is a measure of the goodness of fit of the model used in the analysis. In the present study log likelihood indicate which method that is most likely to predict the data. Comparison of log likelihoods between methods using A- and H-matrices (table 8) showed marginal better likelihood for the method using H-matrix for all traits except total born, where they were the same. This suggests that the ssGBLUP method fits the data better than BLUP method. It agrees with the expectations of more accurate predictions when including G-matrix, and is in accordance with Veerkamp et al. (2011).

Table 9 showed very little change for log likelihood for different weightings, in percentage, of A-matrix into H-matrix. However, it seemed like the likelihood was getting worse when giving more weight to A. Thus, it may be concluded that the G matrix explains all genetic variance and no weighting of the A matrix is needed. For future studies it would be

recommended to do a prediction ability test of the different weightings to obtain more information.

5.3 Predictive ability

Predictive ability was improved (table 10) for the methods including genomic relationship compared to the method using pedigree relationship. This was true for all traits, independent of having many or few genotyped animals. Better predictive ability for the methods using genomic relationship indicates that these methods give better predictions than BLUP. There was neither difference among the methods including genomic relationship, GBLUP and ssGBLUP, nor between methods comparing different variance components (GBLUPg compared to GBLUPa and ssGBLUPg compared to ssGBLUPa).

The fact that predictive ability for total born also was best for the method including genomic relationship, with results of $corr(\hat{y}, y)$=0.28 and 0.23 for BLUP and ssGBLUP respectively, demonstrates that implementing genomic relationship in combination with pedigree relationship gives better predictions even when the main portion of the animals are non-genotyped. Christensen et al. (2012) found similar results when comparing BLUP, GBLUP and original and adjusted single-step methods for genetic gain and feed conversion ratio in Danish Duroc pigs. Overall the results showed highest accuracy for adjusted single-step methods (similar to ssGBLUP in the current study) also when it came to the datasets with many non-genotyped animals. The GBLUP method was not tested for total born in the current study because of very few genotyped- compared to phenotyped animals, but as found by Christensen et al. (2012) there is expected that adjusted single step method would give the best predictions when a dataset contain few genotyped animals.

All regression coefficients deviated from one (table 10), which mean that the predictions are biased. Regression coefficients for observed phenotype on predicted phenotype, $b(y\ on\ \hat{y})$, showed quite similar results for all methods, while there was noteworthy difference in predictive ability between BLUP and the other methods. This indicates that methods including genomic relationship are more accurate, but not less biased. Regression coefficients for observed phenotype adjusted for environment on estimated breeding values, $b((y - \hat{\beta}X)on\ \hat{\mu})$, showed only small differences between the various methods including genomic selection, but the values for the BLUP method were lower than the other methods for growth

and lean meat percentage and higher for feed consumption and total born. It is challenging to explain these differences.

Table 11, 12, 13 and 14 displays correlation between the different methods tested on the estimated breeding values. The results showed high correlation between the GBLUP and ssGBLUP methods and less correlation between BLUP and any of the other methods. This could mean that the methods including genomic relationship predict approximately the same values for GEBVs. A lower correlation between BLUP and the other methods could indicate that the EBVs/GEBVs are unequal. In practical breeding work this would lead to a re-ranking of the animals, which will lead to different selection of breeding candidates. The standard deviation (on diagonal in table 11, 12, 13 and 14) was the lowest when using BLUP compared to the other methods. This agrees with the figures in appendix 2 which showed increased variance in estimated GEBVs/EBVs by ssGBLUP method compared to BLUP.

Predictive ability was measured in two ways, see table 10. The one measuring correlation between EBVs/GEBVs and observed phenotype adjusted for environment, $corr(\hat{\mu}, y - \hat{\beta}X)$ gave the lowest prediction for all traits and methods. This suggests that the environmental effects (fixed-, random- and regression effects) have large effect on phenotype and breeding values. The figures in appendix 2 show that fixed effects have most effect, except for total born where regression effects explain more. Thus, it can be concluded that the difference in predictive ability for the two parameters is not caused by implementation of GS.

6 Conclusion

- The variance components did not change significantly when implementing GS. This indicates that there is unnecessary to estimate variance components based on genomic relationship when implementing GS.
- For future studies it could be performed a comparison of estimated variance components based on different relationship with multi-trait model because the results in the present study may have been influenced by data set and choice of model
- GBLUP and ssGBLUP obtained better likelihood and predictive ability than BLUP
- Regression coefficients demonstrated that all methods were biased
- Choice of method had larger effect when estimating breeding values than estimating variance components.
- Adjusted single step method is a convenient way to implement GS, obtaining better predictions of breeding values for both genotyped and non-genotyped animals.

7 Reference list

Aasmundstad, T. (2014). *Optimal selection for stayability and functionality in pigs*. Ås: Norwegian University of Life Sciences, Animal and Aquacultural Sciences.

Aasmundstad, T., Gjerlaug-Enger, E., Grindflek, E. & Vangen, O. (2014a). Genetic trends of conformation traits and genetic correlations to osteochondrosis in boars. *Animal*, 8 (7): 1045-1052.

Aasmundstad, T., Olsen, D., Sehested, E. & Vangen, O. (2014b). The genetic relationships between conformation assessment of gilts and sow production and longevity. *Livestock Science*, 167: 33-40.

Andersen-Ranberg, I., Martinsen, H. K. & Olsen, D. (2013). *Fokus på tilvekst etter 100 kg*. Husdyrforsøksmøtet Institutt for husdyr- og akvakulturvitenskap, UMB,Norges veterinærhøgskole.

Bourdon, R. M. (2000). *Understanding animal breeding*. 2nd ed. Upper Saddle River, NJ: Prentice Hall. xix, 538.

Christensen, O. F., Madsen, P., Nielsen, B., Ostersen, T. & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal*, 6 (10): 1565-1571.

Clutter, A. C. & Brascamp, E. W. (1998). Genetics of perfomance. . In Rotschild, M. F. & Ruvinsky, A. (eds) *The genetics of the pig*, pp. 427-462. Wallingford, UK: CAB International.

Ducos, A., Bidanel, J. P., Ducrocq, V., Boichard, D. & Groeneveld, E. (1993). Multivariate Restricted Maximum-Likelihood-Estimation of Genetic-Parameters for Growth, Carcass and Meat Quality Traits in French Large White and French Landrace Pigs. *Genetics Selection Evolution*, 25 (5): 475-493.

Forni, S., Aguilar, I. & Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution*, 43 (1): 1.

Gao, H., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S. & Su, G. (2012). Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genetics, Selection, Evolution : GSE*, 44 (1): 8-8.

Gilbert, H., Bidanel, J. P., Gruand, J., Caritez, J. C., Billon, Y., Guillouet, P., Lagant, H., Noblet, J. & Sellier, P. (2007). Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. *J Anim Sci*, 85 (12): 3182-8.

Gjerlaug-Enger, E., Kongsro, J., Ødegård, J., Aass, L. & Vangen, O. (2012). Genetic parameters between slaughter pig efficiency and growth rate of different body tissues estimated by computed tomography in live boars of Landrace and Duroc. *animal*, 6 (01): 9-18.

Grindflek, E. (2013). Norsvin genomics tools - increase genetic gain by 20-40 %. Available at: http://norsvin.no/Media/Files/PDFer-enkeltartikler/Norsvin-Genomics (accessed: 14.01.2014).

Haley, C. S., Avalos, E. & Smith, C. (1988). Selection for litter size in the pig. *Animal breeding abstracts* 56: 319-332 (Sited from Southwood, O.I. & Kennedy, B.W. 1990).

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, 28 (6): 476-490.

Holm, B., Bakken, M., Klemetsdal, G. & Vangen, O. (2004). Genetic correlations between reproduction and production traits in swine. *Journal of Animal Science*, 82 (12): 3458-3464.

Ibáñez-Escriche, N., Forni, S., Noguera, J. L. & Varona, L. (2014). Genomic information in pig breeding: Science meets industry needs. *Livestock Science*, 166 (0): 94-100.

Johnson, R. K., Nielsen, M. K. & Casey, D. S. (1999). Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. *Journal of Animal Science*, 77 (3): 541-557.

Kongsro, J. (2009). *Bruk av datatomograf(CT) i avlsarbeidet på gris-Norsvin Delta*. Husdyrforsøksmøtet 2009, Thon Hotel Arena: Institutt for husdyr- og akvakulturvitenskap, UMB,Norges veterinær høgskole, Veterinærinstituttet. 594 pp.

Legarra, A., Aguilar, I. & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92 (9): 4656-4663.

Lillehammer, M., Meuwissen, T. H. E. & Sonesson, A. K. (2011). Genomic selection for maternal traits in pigs. *Journal of Animal Science*, 89 (12): 3908-3916.

Lopes, M., Silva, F., Harlizius, B., Duijvesteijn, N., Lopes, P., Guimaraes, S. & Knol, E. (2013). Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. *BMC Genetics*, 14 (1): 92.

Madsen, P. & Jensen, J. (2013). *DMU: A Package for Analysing Multivariate Mixed Models. Verison 6, release 5.2*: (available at: http://www.dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf).

Meuwissen, T., Hayes, B. & Goddard, M. (2013). Accelerating Improvement of Livestock with Genomic Selection. *Annual Review of Animal Biosciences, Vol 1*, 1: 221-237.

Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157 (4): 1819-1829.

Nguyen, N. H. & McPhee, C. P. (2005). Genetic parameters and responses of performance and body composition traits in pigs selected for high and low growth rate on a fixed ration over a set time. *Genetics Selection Evolution*, 37 (2): 199-213.

Norsvin. (2013). Annual Report. Hamar.

Ramos, A. M., Crooijmans, R. P. M. A., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., 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).

Rothschild, M. F. & Bidanel, J. P. (1998). Biology and Gentics of Reproduction. In Rothschild, M. F. & Ruvinsky, A. (eds) *The gentics of the pig*, pp. 313-343. Wallingford, UK: CAB International.

SAS. (2013). SAS/CONNECT® 9.4 User's Guide, Second Edition. Cary, NC: SAS Institute Inc.

Sobczyńska, M., Blicharski, T. & Tyra, M. (2013). Relationships between longevity, lifetime productivity, carcass traits and conformation in Polish maternal pig breeds. *Journal of Animal Breeding and Genetics*, 130 (5): 361-371.

Sonesson, A. K., de Greef, K. H. & Meuwissen, T. H. E. (1998). Genetic parameters and trends of meat quality, carcass composition and performance traits in two selected lines of large white pigs. *Livestock Production Science*, 57 (1): 23-32.

Southwood, O. I. & Kennedy, B. W. (1990). Estimation of direct and maternal genetic variance for litter size in Canadian Yorkshire and Landrace swine using an animal model. *J. Anim. Sci.*, 68 (7): 1841-1847.

Su, G. & Madsen, P. (s.a.). *User's guide for gmatrix version2 (available at: http://dmu.agrsci.dk/Gmatrix/Doc/Current/GmatrixV2-User-Guide.pdf)*.

van Wijk, H. J., Arts, D. J., Matthews, J. O., Webster, M., Ducro, B. J. & Knol, E. F. (2005). Genetic parameters for carcass composition and pork quality estimated in a commercial production chain. *J Anim Sci*, 83 (2): 324-33.

Veerkamp, R. F., Mulder, H. A., Thompson, R. & Calus, M. P. (2011). Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *J Dairy Sci*, 94 (8): 4189-97.

Appendix 1

Plots showing the correlation between predicted and observed phenotype, $orr(\hat{y}, y)$, as a measure of predictive ability for growth, feed consumption, lean meat percentage and total born. The plots are presenting the results when using the BLUP and ssGBLUPg methods. The plots for the other methods were very similar to the plots of the ssGBLUP method, and therefore not displayed here.

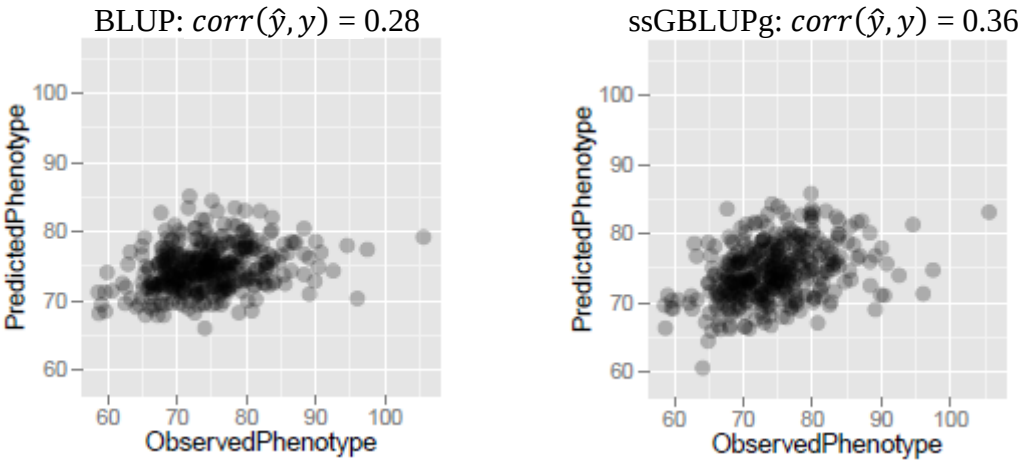


Figure 1: Correlation between predicted and observed phenotype, $corr(\hat{y}, y)$, for the two methods BLUP and ssGBLUP for growth.

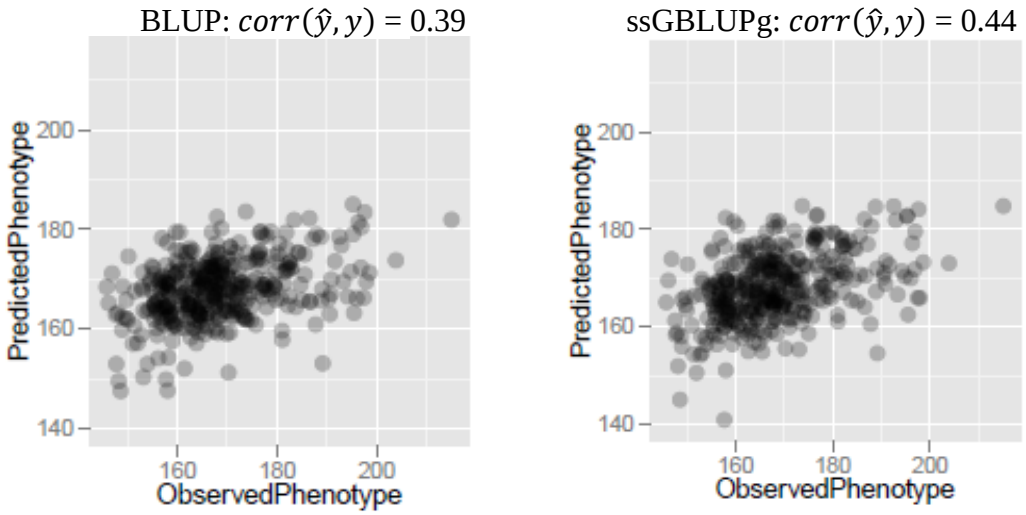


Figure 2: Correlation between predicted and observed phenotype, $corr(\hat{y}, y)$, for the two methods BLUP and ssGBLUP for feed consumption.

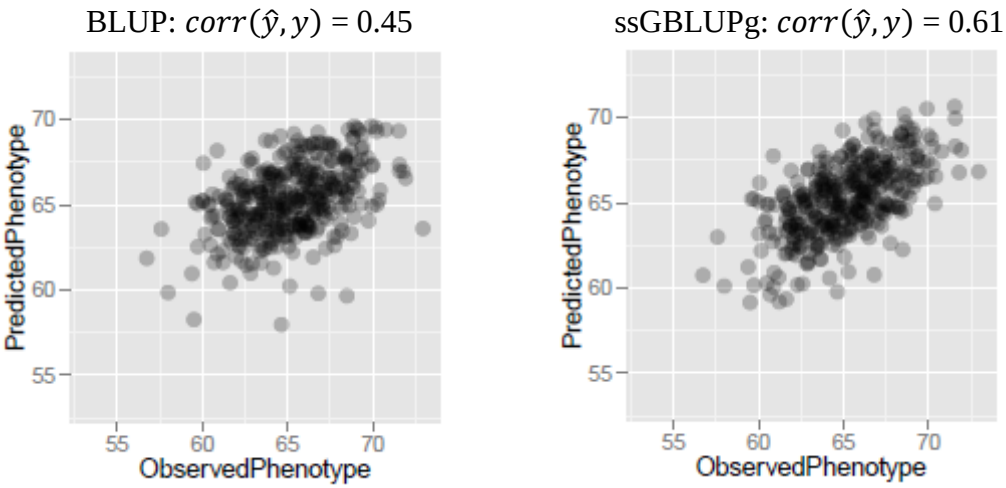


Figure 3: Correlation between predicted and observed phenotype, $corr(\hat{y}, y)$, for the two methods BLUP and ssGBLUP for lean meat percentage.

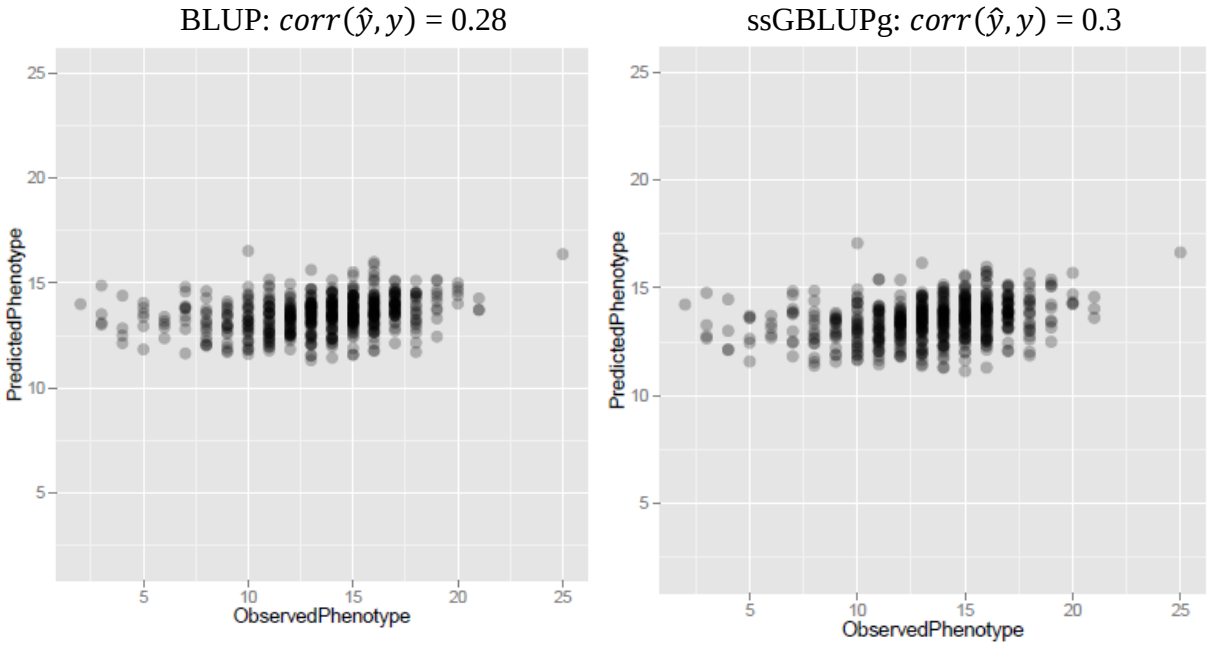


Figure 4: Correlation between predicted and observed phenotype, $corr(\hat{y}, y)$, for the two methods BLUP and ssGBLUP for total born

Appendix 2

Figures presenting the distribution of fixed-, random- and regression effects and EBVs/GEBVs estimated based on relatives for the animals in the validation set for the four traits used in current study. Sum of these values are used to predict phenotype.

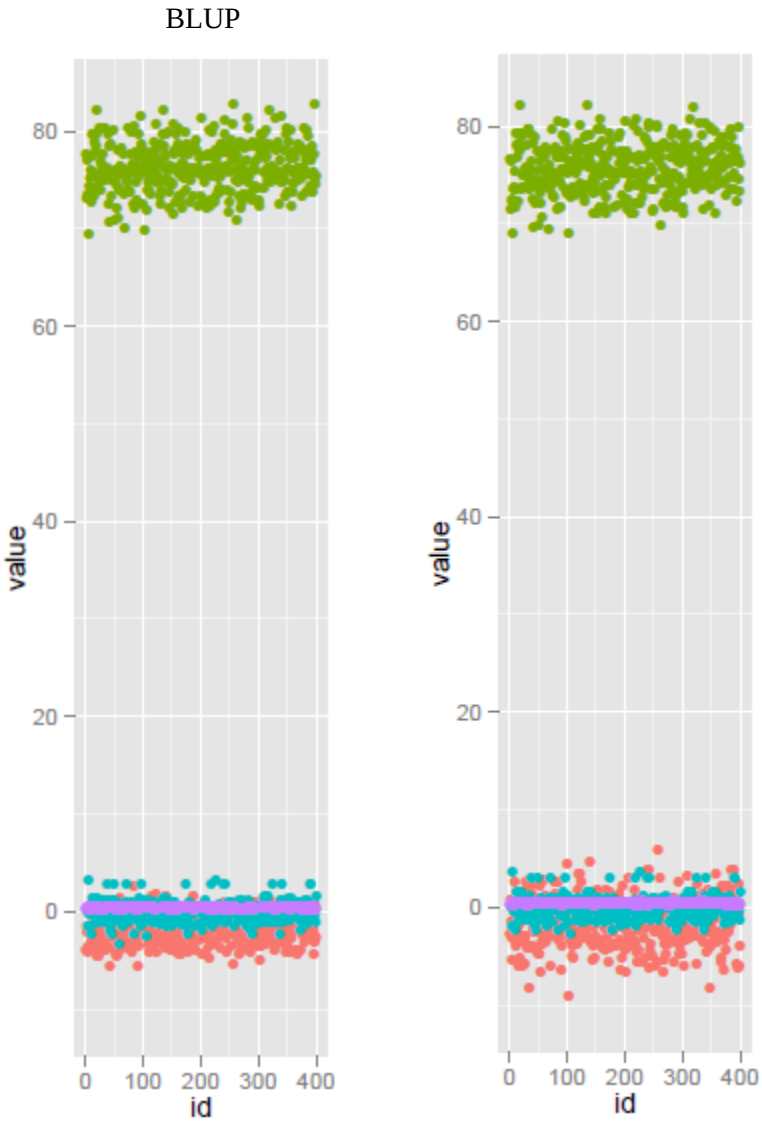


Figure 5: Distribution of random effects (in blue), fixed effects (in green) and regression effects (in purple) and EBV estimated based on relatives for the animals (in red) in the validation set analyzed for growth.

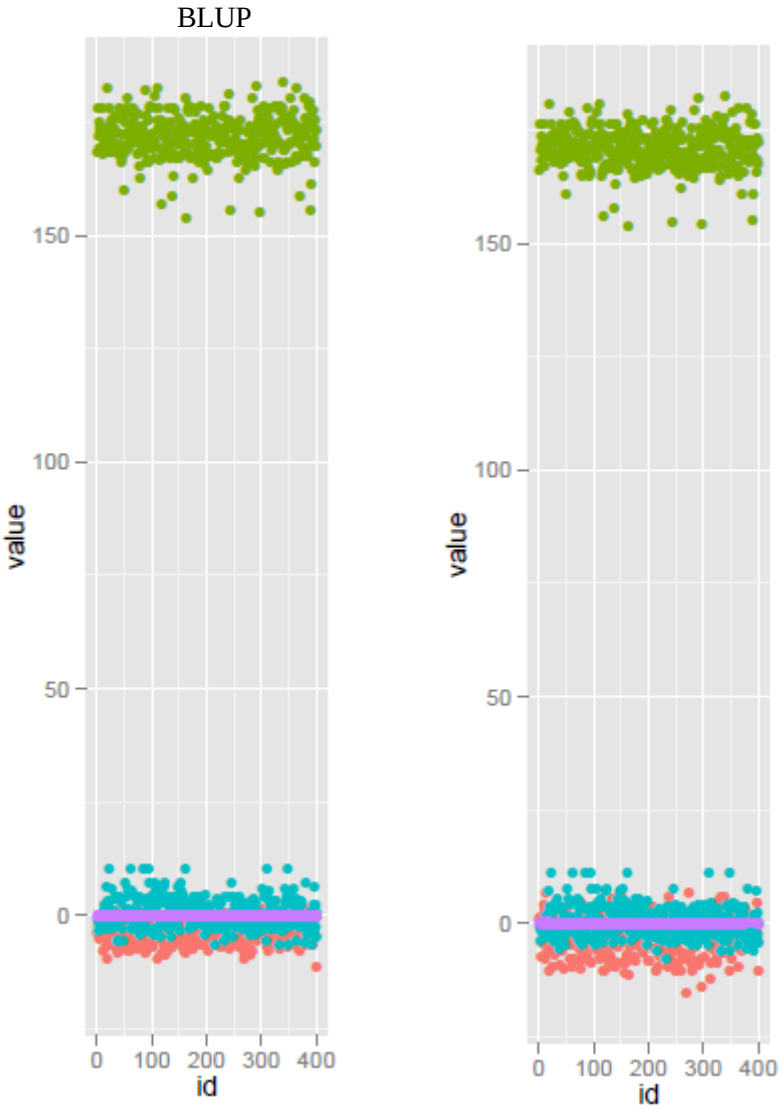


Figure 6: Distribution of random effects (in blue), fixed effects (in green) and regression effects (in purple) and EBV estimated based on relatives for the animals (in red in the validation set analyzed for feed consumption.

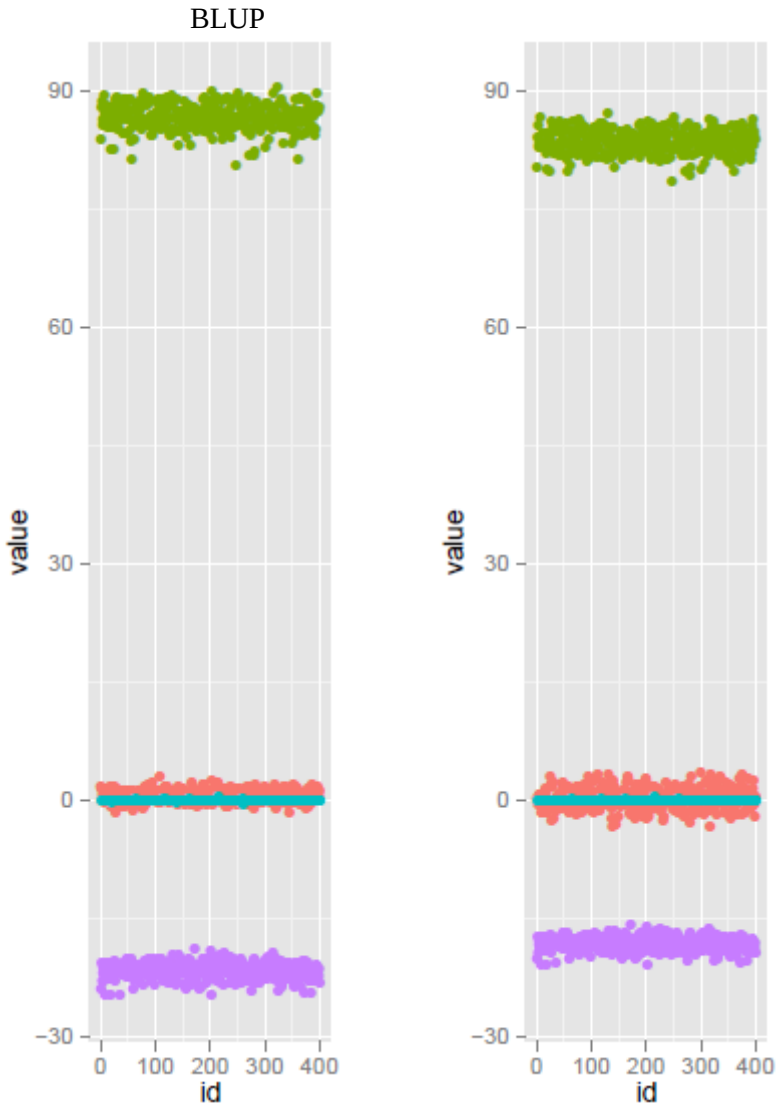


Figure 7: Distribution of random effects (in blue), fixed effects (in green) and regression effects (in purple) and EBV estimated based on relatives for the animals (in red) in the validation set analyzed for lean meat percentage..

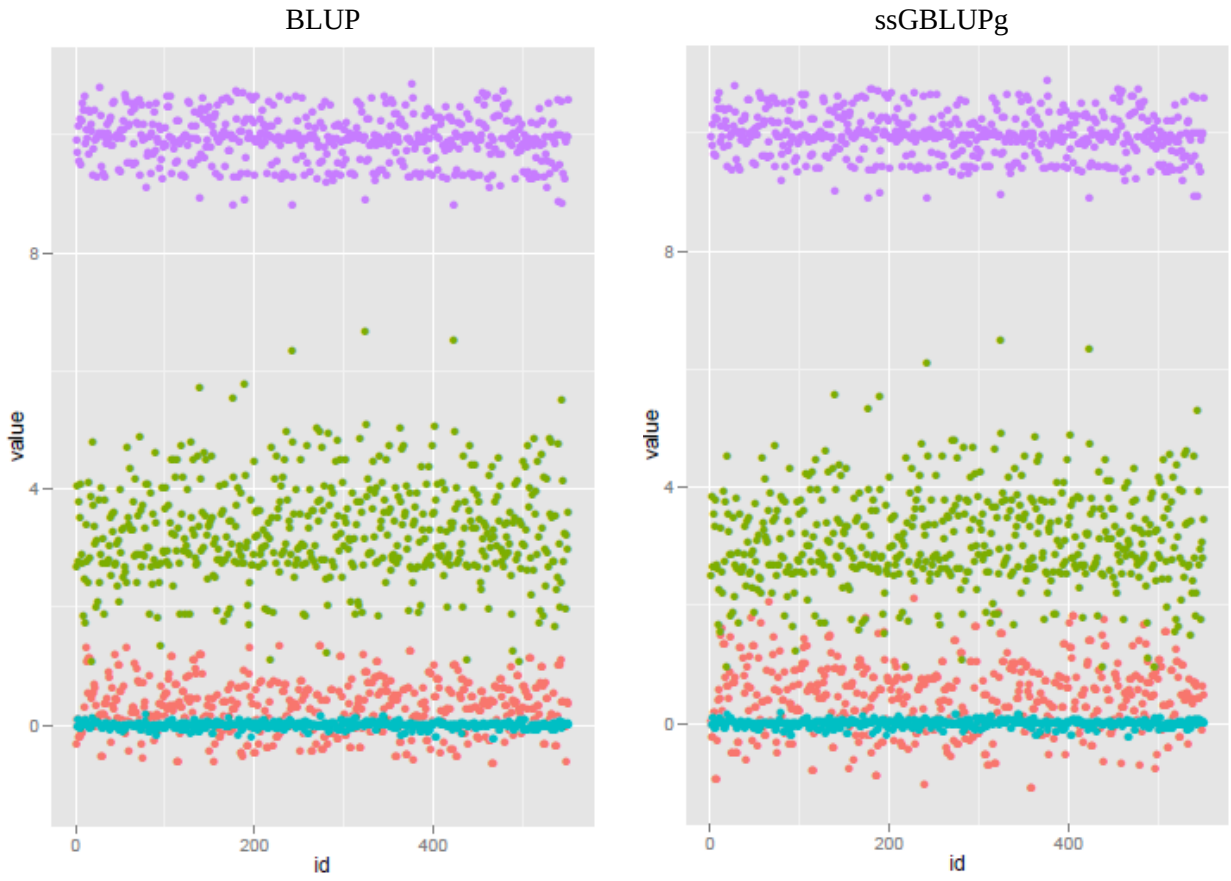
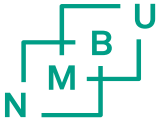


Figure 8: Distribution of random effects (in blue), fixed effects (in green) and regression effects (in purple) and EBV estimated based on relatives for the animals (in red) in the validation set analyzed for total born.



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no