Norwegian University
of Life Sciences

Master's Thesis 2016 30 ECTS
Department of Animal and Aquacultural Science

Genomic predictions including known QTL for reproduction traits in swine

Irene Häfliger
Master in Animal Breeding and Genetics

Master's Thesis

Genomic predictions including known QTL for reproduction traits in swine

Irene Häfliger

Ås, Norway

December 2016

Supervisor

Prof. Theodorus H.E. Meuwissen

Co-supervisor

PhD. Eli Grindflek

Department of Animal and Aquacultural Science

Norwegian University of Life Sciences

Ås, Norway

Acknowledgements

It would not have been possible to write this thesis without the support of many. Therefore, I want to express my appreciation to everybody who made this project possible.

I want to thank my supervisor Theo Meuwissen, who accepted to supervise this thesis. I appreciate his constructive inputs regarding my work.

I want to acknowledge the importance of Eli Grindflek supporting this project. She leaded me through this work as my co-supervisor. I appreciated the discussions with her and her motivating ideas.

Grateful thanks go to the breeding company SUISAG, which supported my thesis financially as well as with data for the analysis. Especially I want to thank Andreas Hofer who acted as my contact and represented my thesis within SUISAG. He defended and supported this project and led a helping hand with good ideas and suggestions.

I would like to send my gratitude to the worker and PhD students at IHA. Namely are those Oscar Iheshiulor, Rajesh Joshi, Yu Xijiang and Maren van Son, who always spontaneously helped me when I needed their help. Thank you very much for your support with instructions, explanations and refreshing discussions.

At last, I also want to thank all my friends and family for their never-ending encouragement and support.

Abstract

Breeding values are the fundament on which the selection of the next generation is based on and should therefore be as reliable as possible. The implementation of marker-assisted selection (MAS), where major genes are included in the breeding value estimation (BVE), was the first method to use genomic information within the BVE. However, with the introduction of genomic selection (GS) it is possible to acknowledge the effects of several thousand markers. In this project, it is explored if the combination of these two genomic methods can improve the prediction accuracy of genomic best linear unbiased prediction (GBLUP).

The study was carried out on the pig breed Swiss Large White. Used were deregressed breeding values of the reproduction traits number of piglets born alive (NBA), the proportion of underweighted piglets (UWP), the survival rate (PS), and the interval weaning to oestrus (IWO). Regarding the inclusion of additional markers, SNP panels in different extents were built based on markers that are known to be of importance from literature. Many QTLs were taken from the pig QTL database to build genomic relationship matrices, one with all QTL-markers (QTL-matrix) and one including only markers that were associated with reproduction traits (rQTL-matrix). They were added to the GBLUP model as additional random effect. Furthermore, few markers associated with the *Escherichia Coli* F4ab/ac (*E.coli*) resistance and the trait number of piglets born alive were added as fixed effects within GBLUP. A cross validation was performed based on the 400 youngest animals in the data set.

Improvements of the likelihoods were observed, when an additional QTL-matrix was included. Significant changes were detected for the trait NBA, by including the rQTL-matrix and for the trait PS, by including the *E.coli*-resistance marker. The prediction accuracy was not improved by giving QTL-markers more weight within the GBLUP model for the data set used.

Even thought, the method did not show any improvement of the prediction ability, the goodness-of-fit improved. In a more powerful data set, the improvements could even enhance the prediction accuracy.

Table of contents

List of tables

List of figures

List of tables in the appendix

List of figures in the appendix

1 Introduction

In European agriculture, pigs are one of the major species in livestock production. According to the Food and Agriculture Organization of the United Nations (FAO) pig farming is important regarding the growing demands for meat around the globe (FAO 2016b). The report of FAO concerning global food markets shows that pork is the most produced meat in the world together with chicken. While the production seems to stagnate, the amount of pork traded is increasing (FAO 2016a). Pigs are especially interesting when it comes to the need for fast-growing and feed-efficient livestock husbandry (FAO 2016b). A tool to achieve more sustainable livestock production is animal breeding. Animal breeding is based on the idea to genetically improve defined traits in a population. Genetically, this means a shift of the allele frequencies from the unpreferable ones to the preferable ones.

In pig breeding, reproduction traits are of major interest. An animal that is not able to give birth is of no interest in a breeding programme that wants to lead the breed in a sustainable and healthy way. Furthermore, the reproduction is a prerequisite for farmers to maintain their livestock and their financial independence. Problems with reproduction can occur before and during pregnancy, as well as while and after giving birth (Velasco 2011). A sow's ability to get pregnant and carry the pregnancy out is very important for an economically oriented farmer. Not only is the number of animals that are born alive important, but also that they are healthy and strong. Large litters often have the disadvantage of lower than desired birth weights, which leads to decreased chances of survival for the piglets. Another factor affecting the ability of piglets to survive is the amount of pathogenic bacteria in the environment. *Escherichia Coli* is one of the most common causes for diarrhoea in piglets and can lead to their death due to dehydration (Nagy & Fekete 2005). Pregnancy has a serious impact and is risky for the mother pig. It demands a lot of energy to carry out and nurse a whole litter of piglets. In order to improve reproduction traits in pigs, breeding values are published, which obtain information to identify the superior animals.

The models used for the breeding value estimation were introduced in the late 20th century (Henderson 1975). However, the goal was always to separate genetic from environmental effects as well as possible and provide accurate criteria for the selection of the next generation. Still, the goal is to improve these models to receive better predictions of how an animal will perform. With increased knowledge about the genome and its functionality, major genes were detected affecting

the fitness or the performance of an animal. Examples of such genes are the halothane marker in pigs, that is associated with meat quality (Henckel et al. 1992) or IGF2, which is associated with muscle growth in pigs (Van Laere et al. 2003). Recently, markers had been found that represent the *E.Coli* resistance (Neuenschwander et al. 2013). These genes can be used for the marker-assisted selection (MAS), where single marker or a small group of markers are added to the breeding value estimation, with regards to increase their accuracy.

MAS was the initiation of a shift in animal and plant breeding from breeding based on pedigree and phenotypic information, to the use of genomic information. However, it was not until recent technological improvements in molecular genetics made it possible to obtain genetic information in form of several thousand single-nucleotide polymorphisms (SNPs) in a cost-efficient manner. Theoretically, with the knowledge about the genome, the whole genetic variance can be explained (Goddard et al. 2010). Therefore, the accuracy of estimated breeding values is improved immensely with the inclusion of genomic information of some thousand SNPs. Breeding values, with which the farmer compares the different animals and decides which one fits best to his breeding goals, had never been as accurate as now. Genomic selection (GS), as it was introduced in 2001 (Meuwissen et al. 2001), is becoming the standard method of selection in all important livestock species. Traditionally, traits are measured when the animal is adult and selection is performed afterwards. With GS the animals forming the next generation can be selected with a higher reliability at an earlier stage, even though proof of their future performance is not yet available (Meuwissen et al. 2001). Considering the big amount of genomic information available, it is vital to figure out what share of it is highly associated with the traits that we want to improve.

Many studies have been conducted to find specific genes or loci that have an effect on a quantitative trait (e.g., (Bergfelder-Druing et al. 2015; Rampoldi et al. 2011). They are called quantitative trait loci (QTL), and are detected if there is a linkage disequilibrium (LD) between a marker and a quantitative trait (Falconer & Mackay 1996). LD occurs when markers are non-randomly associated due to selection, migration, random drift or mutation (Falconer & Mackay 1996). Therefore, the effect of these markers on traits can be estimated and markers with significant effect can be detected. The outcomes of QTL studies depends on the architecture of a trait. The result can range from a few positions that show a major importance to the phenotypic variation to almost no significant position, when many genes are involved in the gene expression.

In the QTL database (Hu et al. 2005) thousands of QTLs are publicly available for the most important farm animals. Currently (accessed 7.9.2016) 16,031 QTLs are registered from over 500 publications for pigs alone (Hu et al. 2016).

1.1 Objectives

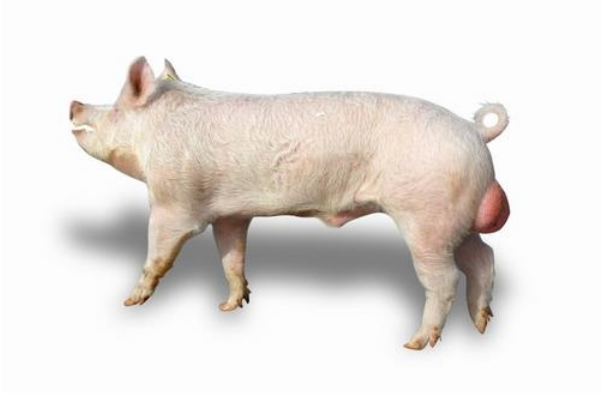
The most common application of GS, the genomic best linear unbiased prediction (GBLUP), is used in this study to calculate genomic estimated breeding values (GEBVs) for reproduction traits in pigs. The breed of interest is the Swiss pig breed Large White and the traits concerned are part of the breeding value estimation for reproduction of the breed itself. The study is conducted on a real-life data set obtained from the Swiss company SUISAG. It is investigated, if the combination of marker assisted selection and genomic selection can improve the accuracy of the GEBVs estimated with GBLUP. More precisely, selected QTL panels from the database (Hu et al. 2005) and a few major genes available in the data set will be added to the breeding value estimation for reproduction traits. The major genes are markers associated with the *E.coli* resistance (Neuenschwander et al. 2013) and litter size (Bergfelder-Druing et al. 2015), whose enhancements are of high medical and economical interest, respectively. Many studies were carried out concerning the improvement of the accuracy by modifying the genomic relationship matrix (G-matrix) (e.g., (dos Santos et al. 2016; Su et al. 2014; Zhang et al. 2010). In this work, the G-matrix will not be modified, but the statistical model is expanded with an extra genomic factor. Therefore, the purpose of this study is to explore if it is beneficial to include genomic information additionally, alongside with the G-matrix, in the model. The goal is to establish the accuracy of the GEBVs and explore whether the accuracy of GEBVs can be improved by adding trait-associated markers to GBLUP model.

2 Background

2.1 Breed description Swiss Large White

The breed Swiss Large White (SLW) originated from small breeds kept in the countryside in Switzerland and was enriched over time with the English Yorkshire breed (SUISAG 2015a). Imports from several countries such as the Netherlands, Germany, England, France and Finland helped to evolve the genetic potential of SLW (SUISAG 2015a). In the year 2002, SUISAG decided to divide the population in a maternal and a paternal line and treated the population as two different breeds ever since. Therefore, the SLW dam line is bred as typical maternal line with outstanding reproduction characteristics. In contrast, the sire line is bred for production and meat quality traits.

In Switzerland, the SLW is known as a robust and highly productive breed. Regarding reproduction performance, the animals in production herds show an average of 12.7 piglets born alive per litter and an average of 2.32 litter per sow and year (Hofer 2016b).



Picture 1: Swiss Large White of SUISAG (http://www.suisag.ch/Zucht/Rassen/Edelschwein/tabid/152/Default.aspx).

Swiss Large White is the most common breed in Switzerland. According to SUISAG's annual report concerning the state and trends in the year 2015, the company had 8604 (174 male and 8530 female) animals in the maternal line registered in their herd book. Additionally there were 454 (249 male and 205 female) registered in the paternal line (SUISAG 2015c).

2.1.1 Breeding programme

In general, the breeding programme of SUISAG incorporates two maternal lines, the Large White and the Landrace, and one paternal line, the Large White sire line. The two dam lines are crossed to produce the hybrid sows (F1-sow) for piglet production. The F1-sow is mated to the paternal line to produce the slaughter pigs (SUISAG 2015b) (Figure A1 in the appendix).

According to SUISAG, reproduction, production and conformation are the main components of the breeding goal, with different weights depending on the breed. For the SLW dam line, the relative weights per genetic standard deviation are 49% for reproduction, 32% for production and 19% for conformation traits (Figure 1) (SUISAG 2016). In comparison, the sire line breeding goal does not include reproduction traits, but includes production traits with 88% and conformation traits with 12% (Figure 1) (SUISAG 2016). Within these three main attributes for the dam lines, SUISAG uses traits that are measured by the farmer themselves, or by technicians in a testing station or on the breeding farm. These traits in the breeding goals are based on requests of farmer, scientific findings, genetic potential and the facilities needed for their implementation in the breeding scheme.

Production includes ten traits that can be combined to four groups. The first group is weight gain (29% in the SLW sire line and 27% in the SLW dam line), which consists of daily weight gain on test in the testing station, daily weight gain on the breeding farm and daily weight gain measured on end products in the slaughterhouse (SUISAG 2016). The second group is meat quantity (16% in the sire line and 10% in the dam line), which consists of loin eye area and the amount of lean meat in the carcass. The third group is concerning meat quality (37% in the sire line and 42% in the dam line), which consist of the amount of intramuscular fat, the pH one hour after slaughtering, the pigment level and drip loss, all measured in the loin (SUISAG 2016). The last group is feed efficiency (18% in the sire line and 21% in the dam line) with only trait feed conversion (SUISAG 2016). An overview of all production traits, comparing the dam and sire line, can be seen in Figure 2. A full list with all relative weights within the complex of production can be found in the appendix Table A1.

The complex of conformation is split into three groups. These three groups imply characteristics for teats, legs and type of the body. The group of traits regarding teats (8% in the SLW sire line and 26% in the dam line) include the number of teats on the left and right side, the number of

inverted nipples and the number of non-functional teats (SUISAG 2016). Type (13% in the sire line and 12% in the dam line) includes only two traits, which are the carcass length and the formation of the loins. The biggest and most important group concerns the legs (79% in the sire line & 62% in the dam line) (SUISAG 2016). It includes traits regarding the rear legs such as knock- or bow-legged, side view angle, angle pastern and size of inner claws. Further are the traits side view angle foreleg, the number of bursas and the gait included (SUISAG 2016). Figure 3 gives an overview of all conformation traits, a full list of all production traits with their relative weights can be found in the appendix Table A2.

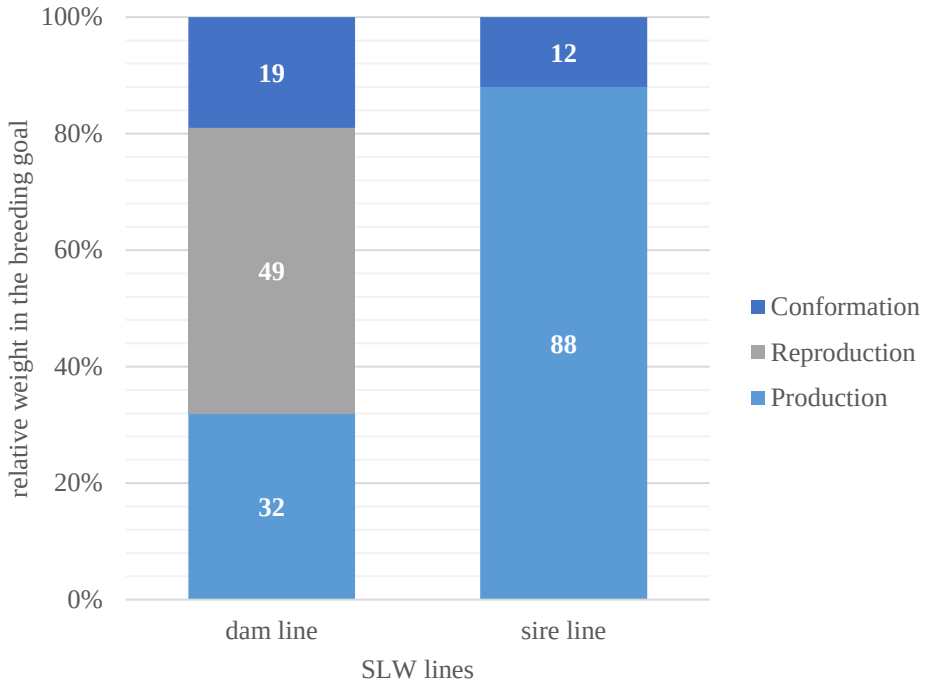


Figure 1: Breeding goals for the Swiss Large White dam line and sire line (SUISAG 2016).

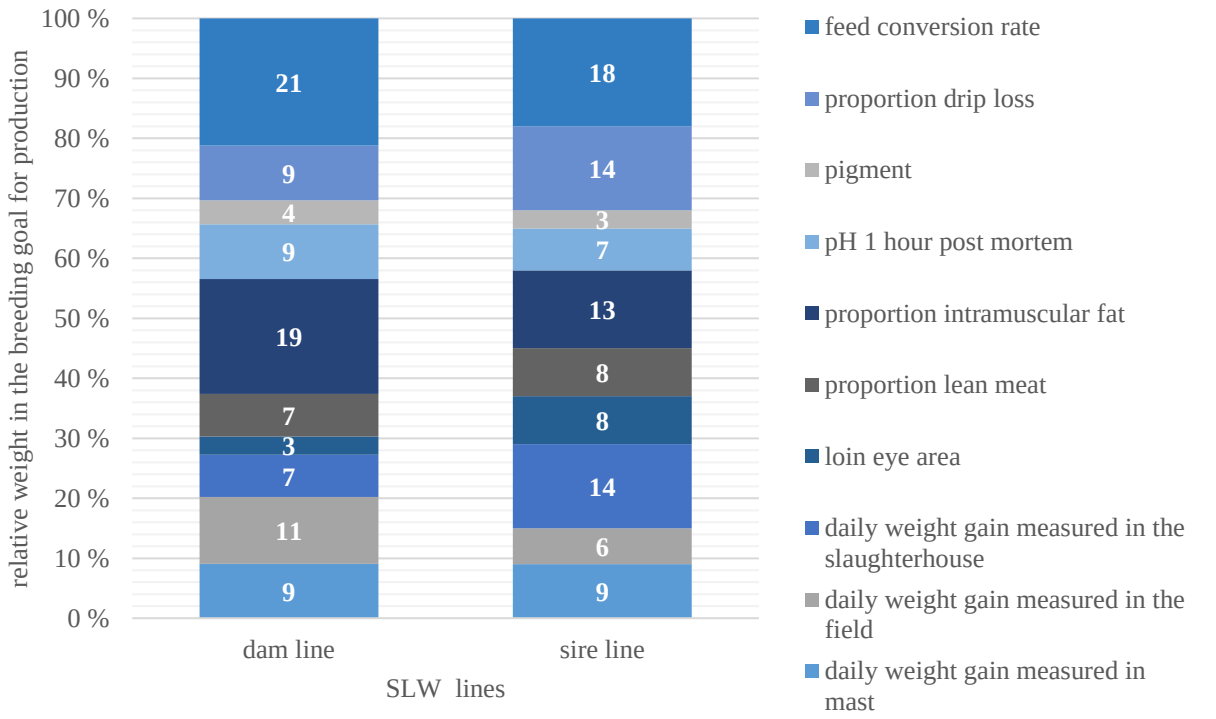


Figure 2 Production traits and their relative weights for the SLW dam line and sire line (SUISAG 2016).

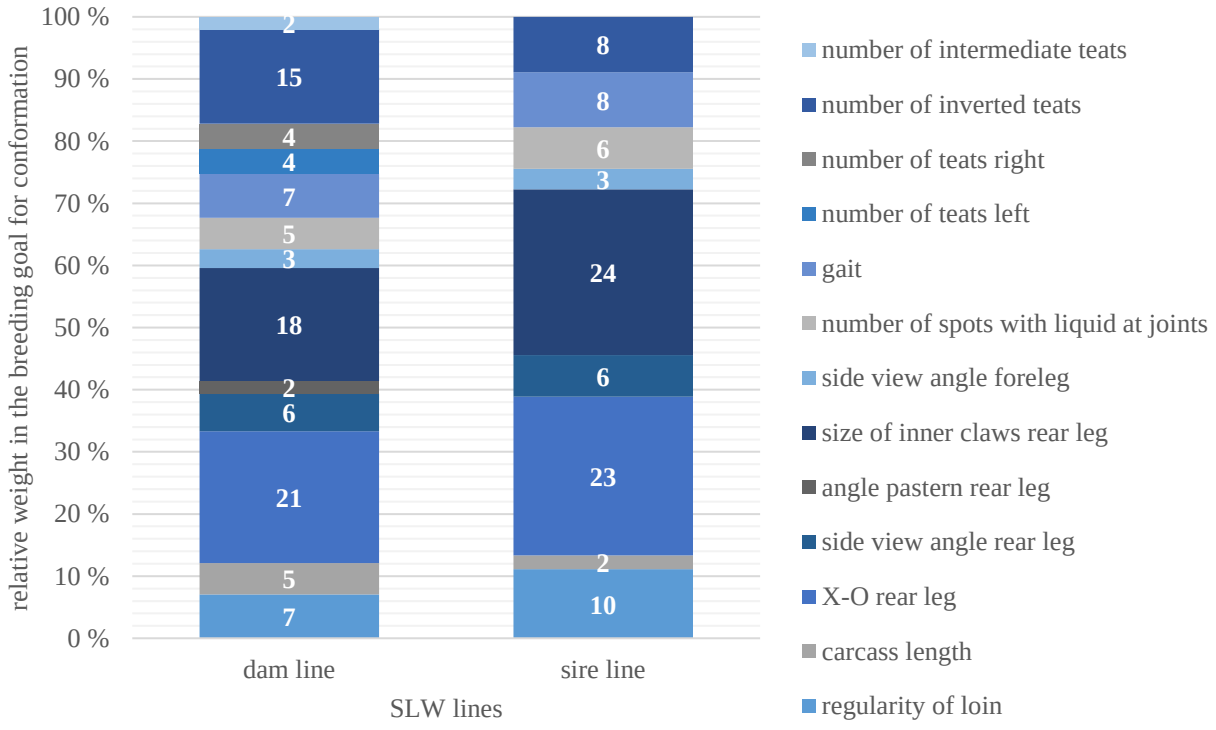


Figure 3: Conformation traits and their relative weights for the SLW dam line and sire line (SUISAG 2016).

2.1.2 Traits of interest: Reproduction

Reproduction traits account for 49% of the breeding goal in the SLW dam line (SUISAG 2016). According to the breeding division of SUISAG, there are four reproduction traits in their breeding programme, which are recorded by the farmers themselves.

- NBA: "number of animals born alive"
  The interest to increase the litter size is obvious, as farmers earn their money per pig sold. Breeding for total number of piglets born indirectly increases the number of piglets born dead. Therefore, the number of animals born alive is a better choice for breeding programmes. The fundament for the inclusion of the trait was the work of Frey (1999).

- UWP: "proportion of piglets born alive below 1kg birth weight"
  This trait is derived from the number of piglets born below 1 kg assessed by the breeder without weighting every single piglet. The trait was introduced in 2012 with regards to improve the birth weight of litters (Hofer, personal communication). From literature it is know that UWP influences the amount of piglets that died (Hellbrugge et al. 2008).

- PS: "the proportion of piglets nursed that are weaned"
  The number of piglets born and the survival rate show an unfavourable correlation, indicating that more piglets die with increased litter size (Hellbrugge et al. 2008). Thus, both traits should be included to make sure that the piglet mortality does not increase uncontrolled. PS was included in 2004 in the breeding goal and is nowadays the most important trait (Hofer, personal communication).

- IWO: "interval weaning to oestrus"
  The interval weaning to oestrus in general is of interest, as it improves the productivity of a sow. The faster a sow returns in the reproduction cycle, the shorter is her unproductive period. Furthermore, is IWO shown to be positively associated with litter size by Wilson and Dewey (1993). Selecting for a shorter interval from weaning to oestrus increases the number of piglets born (Hanenberg et al. 2001). This trait was included in the breeding programme based on the work of Frey (1999).

The relative weights of these traits in the reproduction index are: NBA 30%, UWP 19%, PS 42% and IWO 8% (Figure 4) (SUISAG 2016).

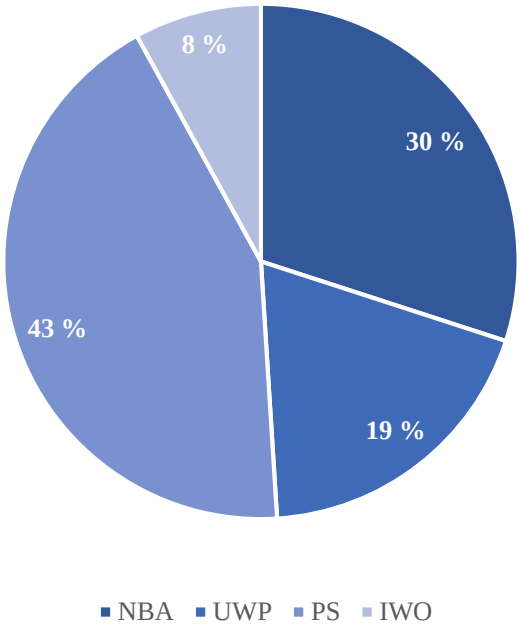


Figure 4: Relative weights of the reproduction trait in the reproduction index regarding the SLW dam line. The traits are number of piglets born alive (NBA), proportion underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO) (SUISAG 2016).

2.2 *Escherichia coli*

Escherichia coli, further called *E.coli*, is one of the most important pathogens causing diarrhoea in piglets (Fairbrother et al. 2005; Schroyen et al. 2012). It is a gram-negative bacterium, which colonises the gut flora naturally, but some pathogenic strains cause diseases. Regarding pigs, the *E.coli* F4 strain is commonly associated with diarrhoea during the suckling period, whereas the *E.coli* F18 strain is associated with post weaning diarrhoea and oedema diseases (Schroyen et al. 2012). The symptoms are caused by the pathogen attaching their fimbriae to a particular receptor in the piglets' small intestine (Nagy & Fekete 2005; Schroyen et al. 2012). After attaching, *E.coli* starts releasing enterotoxins, a protein secrete that prevents enterocytes from executing its absorbing function (Nagy & Fekete 2005). Infections with *E.coli* create high costs in livestock

production due to increased mortality, morbidity, a lower growth rate and medication (Fairbrother et al. 2005). *E.coli* shows more and more resistances to several antimicrobials (Fairbrother et al. 2005) and there is increasing pressure from society to reduce the amount of antibiotics in livestock production. Therefore, there is a great interest to improve the current situation. Since not all piglets show symptoms due to these *E.coli* strains, it has been suspected that there is a genetic resistance in some parts of the pig populations (Meijerink et al. 2000)

2.2.1 Genetics behind *E.coli* resistance

Meijerink et al. (2000) were the first to describe the receptor for *E.coli* F18 in the FUT1 gene and the underlying mutation was found to be *FUT1-c.307A>G*. The single nucleotide polymorphism (SNP) can be used directly to select for *E.coli* F18 resistance, as it is the causative mutation. The F4 fimbriae can be divided into three variants F4ab, F4ac and F4ad, of which the F4ab and F4ac are the most frequent (Fairbrother et al. 2005). It is assumed that the European wild boar is naturally *E.coli* resistant (Jacobsen et al. 2010), but with domestication we seem to have selected against this very useful gene. An explanation is the negative correlation of the *E.coli* resistance with production traits (e.g., weight gain), that we have been artificially selecting for (Nielsen & Johannsen 2004). In 1993, the first association of the *E.coli* F4 resistance with the chromosome 13 was found (Guerin et al. 1993). Ever since many studies were conducted upon the chromosome 13 to find the underlying mutation (e.g., (Jacobsen et al. 2010; Rampoldi et al. 2011; Schroyen et al. 2012). Unfortunately, the fundamental causative mutation, that could be used for all populations has not been located yet (Schroyen et al. 2012).

Nevertheless, several studies claim to have found causative gene (Jacobsen et al. 2010; Jorgensen et al. 2003; Rampoldi et al. 2011). One of them is a Danish group of researcher, that claim to have found the single point mutation *MUC4-g.8227G>C* as causative mutation for resistance against *E.coli* F4ab/ac in a group of animals, a cross between European Wild Boars and Swedish Yorkshire (Jorgensen et al. 2003). This marker is used in the Danish breeding programme (Fredholm 2008). Fredholm (2008) points out that the inheritance is recessive and animals must be homozygote carrier to be resistant against *E.coli* infections.

In 2011, the region of the MUC4 gene on chromosome 13 was refined and six SNPs in complete linkage disequilibrium with the resistance to F4ab/ac located (Rampoldi et al. 2011). This study

had been conducted upon the Swiss Large White breed and the markers found were *ALGA0072075*, *ALGA0106330*, *MUC13-226*, *MUC13-813*, *DIA0000584* and *MARC0006918* (Rampoldi et al. 2011). In a further study the markers *ALGA0106330* and *ALGA0072075* showed a 100% conformity with the *E.coli* F4 resistance (Neuenschwander et al. 2013).

2.2.2 *E.coli* resistance in breeding programmes

The Danish pig research centre (VSP – Videncenter for Svineproduktion) have been working on increasing the frequency of the *E.coli* F4ab/bc resistance gene in their Landrace, Duroc and Yorkshire population since 2003. According to their publications in DanAvl Magasinet, there is a tendency that resistant piglets have a higher growth rate from birth to 30 kg weight than non-resistant piglets (Nielsen & Johannsen 2005). Litter size is revealed to be higher in resistant animals for the breed Yorkshire but not for Landrace (Nielsen & Johannsen 2004). On the other hand, it is known that the *E.coli* F4 resistance is negatively correlated with the portion lean meat, daily growth from 30 to 100 kg, feed efficiency and slaughter losses (Nielsen & Johannsen 2004). After several years of breeding efforts, DanAvl revealed that they completed the F4 breeding project, as the resistance gene is almost fixed in all populations (Nielsen & Svensmark 2010). As a result, the amount of piglets dying due to diarrhoea within the first 5 days after birth has decreased (Figure 5) and the growth between 0 kg up to 30 kg weight has increased (Nielsen & Svensmark 2010).

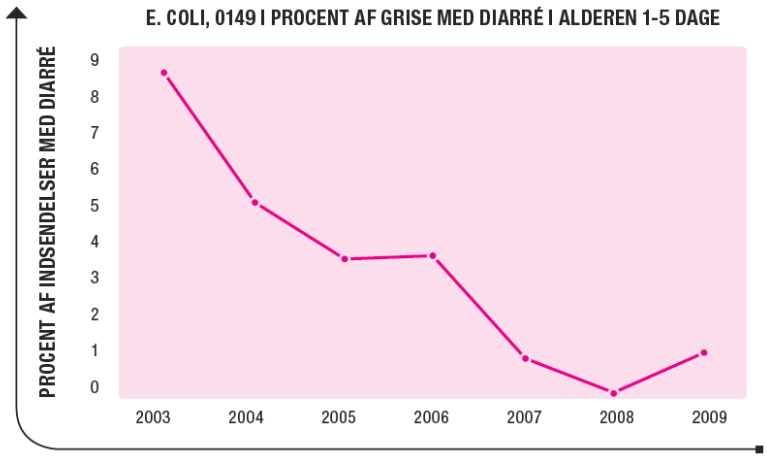


Figure 5: Changes in the proportion of piglets that died because of diarrhoea within the first 5 days after birth in the Danish pig population maintained by DanAvl (Nielsen & Svensmark 2010).

In Switzerland the Large White dam line is resistant to the *E.coli* F18 strain, while the selection is still on-going in the sire line (Luther 2015). They had a rigid selection programme for the marker on the *FUT1* gene found by Meijerink et al. (2000). *E.coli* F4 is now a bigger problem in Switzerland than *E.coli* F18 had been. Its resistance is going to be part of the breeding programme soon. Genotyping key animals for the markers associated with the *E.coli* F4 resistance has started in order to prepare for the selection (Hofer, personal communication). SUISAG plans to improve the level of resistance in the sire line first. The idea behind it is, to use the benefit of having non-resistant sows nursing resistant piglets. With the colostrum, non-resistant sows transfer valuable antibodies, what helps the piglets to compete better with *E.coli*. The odds are high to improve the allele frequency for the resistance gene in near future.

2.3 Marker assisted selection

Traditional breeding values are calculated using phenotypes and animal relationships based on a pedigree. The commonly used method is the best linear unbiased prediction (BLUP) (Henderson 1975). The linear model assumed in BLUP is

$$y = X\beta + Z\,u + e\ ,$$

with y as a $n\,x\,1$ vector of observations. X and Z are known design matrices linking the observations to the fixed regression parameters β and the random effect u, respectively. u is random vector including all animals in the pedigree and has assumed variance $NID(0, A\,\sigma_a^2)$, whereas e is a random vector of size $n\,x\,1$ with the assumed variance and $NID(0, R\,\sigma_e^2)$, respectively (Henderson 1975). The matrix A is the numerator relationship matrix, representing the coefficients of the additive genetic relationship among all animals in a pedigree and R is a $n\,x\,n$ matrix. Further is σ_a^2 the additive genetic variance and σ_e^2 the residual variance. When we calculate breeding values, we are looking for the variable u.

With improvements in molecular genetics, single or group of genes can be detected that affect important traits in the breeding goal or diseases. It can either be direct genes or representative markers obtained by linkage equilibrium or linkage disequilibrium with QTLs. An example used

in breeding programmes is the *DGAT1* gene affecting the milk-fat content in dairy (Grisart et al. 2002). Another example is the halothane gene (*HAL*) (Fujii et al. 1991) in pigs, which affects the meat quality. The *HAL* gene is known to be associated with the expression of the Porcine Stress Syndrome (PSS). PSS is one of the major reason for increased pale, soft and exudative meat (PSE meat) (Jermiah et al. 1999). Carrier of the HAL gene have generally poorer meat quality than non-carrier. In pigs, the *IGF2* gen (insulin-like growth factor 2) was detected with a mutation, affecting the QTL for muscle growth (Van Laere et al. 2003). The Q allele was identified as the stimulating allele for the formation of muscle tissues. The development is especially enhanced during the embryonic phase. Animals inheriting it from their sires have a threefold higher *IGF2* messenger RNA expression in postnatal muscles (Van Laere et al. 2003).

Marker assisted selection (MAS) is the inclusion of single markers or group of markers in the model of the breeding value calculation. It extends the traditional BLUP model with an additional effect. The resulting model is

$$y = X\beta + XMg + Z\,u + e,$$

with mostly the same model terms as above. The term defining the additional effect is X, a design matrix linking the observations to the marker effects, M, a $n\,x\,m$ matrix with the m marker genotypes coded as 0, 1 or 2, depending on the number of minor alleles (0 0, 0 1 or 1 1) and g, a vector of the allele substitution effects for each marker m. The improvement of MA-BLUP compared to traditional BLUP was confirmed by different research groups (Lande & Thompson 1990; Meuwissen & Goddard 1996). However, its success is mainly dependent on the fraction of additive genetic variance explained by the markers used (Lande & Thompson 1990). Lande and Thompson (1990) showed that MAS is more efficient for traits with a low heritability. Similar results were detected by Meuwissen and Goddard (1996). Furthermore, they predicted a higher additional increase in genetic gain for traits that are recorded after selection (Meuwissen & Goddard 1996).

2.4 Genomic Selection

Genomic selection (GS) (Meuwissen et al. 2001) is applied, since technological improvements made it possible to receive genomic data in a bigger quantity to an affordable price. Especially the inventions of modern computer technology and high throughput sequencing are considered as milestones, opening new prospects for modern bioinformatics. Regarding breeding programmes, this development means that we have the possibility to include genomic information into the estimation of breeding values (Meuwissen et al. 2001). The markers used are single nucleotide polymorphisms (SNPs), representing genetic variation in a single base of the genome. Different sizes of SNP-panels are available, but commonly used are panels of 50,000 to 80,000 SNPs. For Pigs, the most common chips are the Illumina Porcine SNP60 BeadChip (Illumina, San Diego, CA, USA) and the GeneSeek custom 80K SNP chip (Lincoln, NE, USA). The difference between MAS und GS is the amount of genomic information used in the model and their relative weight. MAS only uses one or a few genes, whereas GS includes several thousand SNPs at once. Thus, the few marker in MAS receive more weight in the model, while the markers in GS receive all the same and only small weight. The model proposed for GS is

$$y = 1_n \mu + \sum_i M_i g_i + e ,$$

with y containing the observations corrected for fixed effects in a $n\,x\,1$ vector, 1_n as a $n\,x\,n$ identity matrix, μ being the mean over all performances, $\sum_i M_i g_i$ as the sum of all products between marker M_i and its effect g_i at position i and e as an unknown random vector of length n with assumption $NID(0, \sigma_e^2)$ (Meuwissen et al. 2001). This model is called the marker based best linear unbiased prediction (mBLUP). Because of this model includes all markers as regression parameters on the phenotype, the effect of each marker is estimated based on the fraction of animals with genotypic and phenotypic information. Furthermore, we can calculate the predictions for animals that do not have own phenotypes, but do have genomic information available. Meuwissen et al. (2001) showed that these predictions have a higher reliability than traditional breeding values. As a matter of fact, the selection of the next generation is possible at an earlier stage and with a higher accuracy than with the traditional methods (Meuwissen et al. 2001). A stronger selection at an earlier stage can reduce the costs for the raising and performance testing of selection candidates (Meuwissen et al. 2001). Therefore, the introduction of GS can improve the economy of a breeding programme directly. The major benefits for cattle breeder are

the shortage in the generation interval and the improved genetic gain. The pig breeder on the other hand gain most due to the higher accuracy of GS. In fish production, the genomic information is amongst others used to distinguish between individuals, as they have big groups of full sibs.

The marker effects have to be estimated on a training population (Meuwissen et al. 2013). Usually, it includes animals with genomic and phenotypic information that are born within a defined period of time (e.g., the last 5 years). The training population needs to be updated frequently, due to the changing marker frequencies in the new generations and their effects accordingly.

There are two equivalent genomic BLUP models. Namely the marker based model (mBLUP) (Meuwissen et al. 2001) and the model based on a genomic relationship matrix (G-matrix) (GBLUP) (e.g., (Habier et al. 2007). The mBLUP model is computationally more demanding, as it calculates the effect of each marker included in the model, whereas the GBLUP estimates the effects of each animal (Meuwissen et al. 2013). The difference is especially obvious, when few animals have many marker genotypes.

GBLUP is an animal model with a genomic relationship matrix (G) instead of the pedigree based numerator relationship matrix (A). The GBLUP model is

$$y = 1_n \mu + Zg + e,$$

with the same parameters as the mBLUP, simply that Z is a design matrix to link the observations to the animals and g is a vector of random effects with assumption $NID(0, G\ \sigma_a^2)$ and G being the genomic relationship matrix (Habier et al. 2007). In GBLUP the relationship matrix is calculated according to the SNP markers that are identical by state (Habier et al. 2007). A major advantage of the G-matrix compared to the A-matrix is that the relationship between two animals is more accurate, as the pedigree-based relationships are merely expectations (Meuwissen et al. 2013). For instance, in the A-matrix the relationship of full sibs is always ½ but in reality it might range from 0 to 1 (Meuwissen et al. 2001).

2.4.1 Non-linear models

Both genomic BLUP models introduced assume that all markers explain equal amount of the total genetic variance and are therefore linear models. Non-linear models assume a prior distribution and consequently an unequal amount of genetic variance explained by different markers

(Meuwissen et al. 2001). Thus Bayesian models, such as BayesA and BayesianLasso (Legarra et al. 2011), which assume a different amount of variance explained by the SNPS but that all SNPs have an effect, whereas BayesB and BayesC assume that a few SNPs have high effect and the rest has no effect at all. Often the Bayesian methods perform better in simulation studies, but not in real-life data analyses (Meuwissen et al. 2013). However, the most common used method in practice is the GBLUP, which is relatively simple to implement in an already existing breeding programme. It is a computationally stable method, as we do not increase the complexity of the model, but replace the A-matrix with the G-matrix (Meuwissen et al. 2013). Furthermore, it is very attractive for populations with an incomplete or unknown pedigree, considering we are only interested in the genomic relationships (Goddard et al. 2010).

2.4.2 Multi-trait genomic selection

Multi-trait genomic selection (MT-GS) includes several traits in one model. The benefit of multi-trait models is that they adjust for interactions between traits used in a breeding programme (Jia & Jannink 2012). A model with more traits is supposed to include more information and accordingly is more accurate (Jia & Jannink 2012). MT-GS is computationally more demanding than single-trait models. Nevertheless, its advantages are that traits with a low heritability can benefit from high-heritable traits, when they have genetic correlations (Calus & Veerkamp 2011).

2.4.3 Genomic selection in pig breeding

In pig breeding, the most important selection decision is taken when the selection candidates reach the usual slaughter weight. Many of the traits selected for, such as product quality and quantity, are not measured on the living animal itself (e.g., slaughter losses, meat tenderness). They are mostly measured on full sibs. Furthermore, it is difficult to select for maternal traits that can only be observed on female relatives after their first litter (Lillehammer et al. 2011). To use the effect of heterosis, most breeding programmes keep several purebred lines to produce crossbred offspring (Jonas & de Koning 2015). Hence, there is a potential for a large amount of information originating from crossbred animals to be used in the selection of the purebred lines (Jonas & de Koning 2015).

The analysis with genomic data enable the possibility to compare the markers of the selection candidate with the markers of the performing full sibs. It has been proven on real-life data that GS can improve the selection accuracy for traits that are recorded on relatives (Nordbø et al. 2014). According to Lillehammer et al. (2011), GS can improve genetic gain for maternal traits even though these traits have low heritability, are not measured on the selection candidate and cannot be recorded before the first litter. In a further study, Lillehammer et al. (2013) concluded that including female genotypes is beneficial for maternal traits in a breeding programme, when maternal traits are prioritised. Thus, not only the male selection candidates should be genotyped, also the sows from the nucleus farms. For the genomic predictions of crossbred animals, more extensive use of crossbred genotypes in GS is beneficial (Hidalgo et al. 2015).

2.5 Methods used for the data preparation

2.5.1 Imputation

What we call imputation is the procedure of estimating genomic information for missing genotypes, based on the genomic information obtained by a set of genotypes. Usually this is genomic information from the population of interest. Missing genotypes can occur due to technical problems in the laboratory or bad sample quality. Imputation can also be used to implement genomic information for animals that are not genotyped. This has a high accuracy, as long as the individuals that are included in the training data have a close relationship (Pimentel et al. 2013). The accuracy of genomic predictions increases, the more genotypic information included, even if the animals are imputed (Pimentel et al. 2013). Many imputation programmes available (e.g. FImpute, findhap) use pedigree information in addition to the genotypes. These programmes impute first based on the most probable genotype inherited by the parents and then take the most probable allele according to the population into account. Imputation is based on haplotypes, rather than single markers, and is therefore highly accurate (Browning & Browning 2009). The more animals the reference panel includes, the higher is the imputation accuracy. However, imputation can lead to over and underestimated breeding values for extremely bad and good animals, respectively (Pimentel et al. 2015). The reason for this is that the imputed marker

represent the population mean. Hence, inferior animals receive over-estimated marker effects and superior performing animals receive under-estimated marker effects.

2.5.2 Deregression of breeding values

Usually, the estimation of breeding values is based on animals with different amount of information from different origin. For instance, some animals have repeated measurements (e.g., sows with several litters) or information of ancestors (e.g., litters of the mother). By using estimated breeding values (EBVs) in a genomic model directly, we bias the estimation of the genetic effects, as we include information of different sources and amounts. As a matter of fact, EBVs lead to double-counting, when both the offspring and ancestors are genotyped (Ostersen et al. 2011). Due to the negative correlations of prediction errors with breeding values, the inclusion of direct EBVs result in a shrinking of the total genetic variance (Garrick et al. 2009). Furthermore, the different amounts of information used in BLUP shrinks the EBVs according to their reliability. Therefore, deregressed breeding values (drEBV) are calculated and used for the genomic prediction. The method from Garrick et al. (2009) suggests to correct for parent average and the shrinking. What is called shrinking is the circumstance that BLUP narrows the estimates towards the population mean (Garrick et al. 2009). The deregressed observations merely account for the own performance and the offspring performances. However, the drEBVs have heterogeneous variances if the reliabilities of the underlying breeding values are varying between animals (Garrick et al. 2009). Hence, the weights account for the repeated measurements on an individual. The drEBVs with their weights can be used in GBLUP in a weighted analysis directly.

Guo et al. (2010) analysed the use of deregressed breeding values on a simulated cattle population. Their results indicate slightly better performance of the EBVs over the drEBVs. Nevertheless, Ostersen et al. (2011) increased the reliability of GEBVs with 15-39% by replacing the EBVs with drEBVs in a pure-bred pig population. They explain the differences, with the lower double-counting in the cattle data, whereas the pig population showed inferior heritabilities and a high amount of double-counting (Ostersen et al. 2011).

3 Material and Methods

3.1 Data

The Swiss company SUISAG provided the data analysed in this project. The data set includes information of both the maternal and the paternal line of the pig breed Swiss Large White (SLW), with a total of 2,486 animals. Both lines are part of the analysis. 1,911 females and 575 males are available, from which 2,305 animals belong to the maternal line and 171 animals to the paternal line (Table 1). All animals have phenotypic as well as genomic data available. No additional animals will be included in the analysis.

Table 1: Data distribution according to breed and sex for the pig breed SLW.

		breed		
		dam line	sire line	sum
sex	male	428	147	575
	female	1,877	34	1,911
	sum	2,305	171	2,486

The pedigree encompasses 8,211 individuals, of which 2,486 have records and the remaining 5,725 individuals are their ancestors. It reaches back for a maximum of ten generations. The oldest ancestor included was born in 1983. Even though, the paternal and maternal line of SLW are kept as two independent breeds, they share common ancestors within the pedigree.

3.1.1 Phenotypic data

The phenotypic data includes the four reproduction traits: number of piglets born alive (NBA), proportion of piglets born under 1kg birth weight (UWP), proportion of piglets nursed that are weaned (PS), and the interval from weaning to oestrus (IWO). More information about the traits can be found under paragraph *2.1.2 Traits of interest: Reproduction*. The phenotypic information was received as deregressed breeding values and their corresponding weights, which are calculated according to the method described by Garrick et al. (2009). A whole table with

statistical reference numbers of the EBVs, drEBVs and their weights is in the appendix (Table A3).

3.1.2 Genomic data

Originally, the data included information of the 2,486 animals from four different types of marker arrays. They were distributed as 29 genotypes with the Illumina Porcine SNP60 Beadchip version 1 (Illumina, San Diego, CA, USA), 1,779 genotypes with the same chip but version 2, 555 genotypes with the GeenSeek 80K (Gene Seek Inc, Lincoln, NE, USA) and 60 genotypes with a custom chip based on Illumina Porcine SNP60 Beadchip version 2. The data was delivered as imputed genotypes. All genotypes had been imputed for all markers with the programme FImpute. (Sargolzaei et al. 2014). It was decided to restrict the markers used in the analysis, to the SNPs that occur on both the Illumina Beadchip version 2 and the GeenSeek 80K chip. Furthermore, to secure an acceptable level of accuracy for the imputed genotypes, the markers were filtered for a minor allele frequency of 0.01 and a call rate of minimal 0.05. Therefore, 34,879 SNPs build the basis of the whole analysis.

3.1.3 Genomic data regarding the additional effect

To model the supplementary effect, all 16,031 QTLs known and publicly registered on the QTL database for pigs (Hu et al. 2005; PigQTLdb 2016) were taken into account. According to the information available about their flanking markers, two SNP-panels were built. All flanking markers that occurred in the genotypic data described above build the first QTL panel. It consists of 4,205 SNPs. The second panel, with markers specifically associated with reproduction, consists of 1,103 SNPs (Table 2) Table A4 in the appendix shows the distribution concerning the markers across the chromosome for both panels.

Additionally, the markers found by Neuenschwander et al. (2013) were taken into account (Table 2). These SNPs showed complete linkage disequilibrium and 100% conformity with the resistance gene responsible for *Escherichia coli* F4ab/F4ac resistance in SLW (Rampoldi et al. 2011). From the two SNPs found, one was available in the genomic data. This marker was *ALGA0106330*, which is located on chromosome 13. With one marker, the effect modelled has two different effect

levels. Information regarding the marker position and frequency can be seen in Table A5 in the appendix.

Another group of SNPs important regarding reproduction traits were found to be significant to the number of piglets born alive by Bergfelder-Druing et al. (2015) (Table 2). This study was conducted on the breeds Large White and Landrace from Austria, Germany and Switzerland. A genome wide association study was conducted and revealed thirteen significant SNPs affecting litter size in Large White. The intersection of the associated markers found in the work from Bergfelder-Druing et al. (2015) and the available data, resulted in the two markers *MARC0043480* and *MARC0006510*. The first marker is located on chromosome 10 and the second on chromosome 11 (Table A5). As a result, a panel of two markers arose for the last analysis. Two markers can build four different haplotypes, of which all occur in the population. Information regarding the marker positions and frequencies are stated in Table A5 in the appendix.

Table 2: Overview of the marker panels used regarding the additional effects, the number of SNPs included and their use in the single-trait GBLUP.

description	short name	Number of SNPs	effect	sort of factor
all QTLs	QTL	4,205	Random	QTL-matrix
reproduction QTLs	rQTL	1,103	Random	QTL-matrix
NBA marker[1]	LS[1]	2	Fixed	Haplotype
E.coli marker	*E.coli*	1	Fixed	Single marker

[1] *to avoid confusion between the trait NBA and NBA associated markers the markers will further be called litter size markers (LS)*

3.1.4 Reference and validation group

In order to receive comparable output, the 400 youngest animals in the data set were masked. It resembles 16% of the animals with genomic and phenotypic information available and includes animals going back to be born in the middle of 2012. In this validation group were 317 females and 83 males, with 361 belonging to the dam line and 39 to the sire line. All animals have phenotypic information, either due to own performance or performance of progenies. Some statistical reference numbers are indicated in Table 3 regarding the reliabilities of the breeding values, deregressed breeding values and their weights for the reference and the validation group.

Table 3: Statistical reference numbers (minimum – Min, maximum – Max, average – Mean and standard deviation – SD) of the reliabilities of the breeding values (r^2), the reliabilities of the deregressed breeding values (r^2 drEBV) and the weights of the deregressed breeding values (weight) of SLW. The data is divided into the reference group (reference) and the validation group (validation) and further for each trait.

		reference			validation		
		r^2	r^2 drEBV	weight	r^2	r^2 drEBV	weight
Trait[¥]	N	2,086	2,086	2,086	400	400	400
NBA	min	0.38	0.20	1.82	0.42	0.28	2.76
	max	0.99	0.99	67.39	0.97	0.97	54.40
	mean	0.67	0.58	12.31	0.65	0.54	9.77
	SD	0.11	0.16	12.48	0.10	0.14	8.54
UWP	min	0.30	0.20	2.31	0.38	0.24	3.00
	max	0.98	0.98	82.85	0.96	0.96	68.32
	mean	0.64	0.54	13.78	0.63	0.51	11.57
	SD	0.11	0.16	13.84	0.10	0.14	10.32
PS	min	0.32	0.18	3.27	0.34	0.22	4.04
	max	0.99	0.99	132.06	0.95	0.95	95.52
	mean	0.62	0.50	19.72	0.59	0.46	14.39
	SD	0.13	0.18	22.73	0.10	0.15	13.84
IWO	min	0.27	0.13	1.01	0.32	0.18	1.55
	max	0.98	0.98	58.92	0.95	0.94	44.11
	mean	0.54	0.40	6.74	0.54	0.38	5.57
	SD	0.14	0.20	9.24	0.12	0.17	6.50

[¥] *The traits are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO)*

3.1.5 Descriptive statistic of the relationships

The relationship according to the G-matrix and the A-matrix, are specified for all animals with records (Table 4). The calculation of the G-matrix is explained as part of the paragraph *3.2.2.1*

Genomic breeding values. The first part of Table 4 shows the relationship between the different animals (off-diagonal elements) and the second one shows the within animal relationship (diagonal elements). The within animal relationship is related to the inbreeding in the population. Generally, the genomic relationship has a distribution with mean zero and a higher variance as the pedigree based relationships, as more information is included. According to the pedigree-based relationship, the validation group has a slightly higher average of both between animal and within animal relationship. The same can be observed for the genomic relationship between animals, but not within animals (Table 4).

Table 4: Statistical reference numbers (minimum – Min, maximum – Max, average – Mean and standard deviation – SD) of genomic and pedigree based relationship of SLW within the reference group (ref), within the validation group (Val) and between the reference and validation group (between ref and val) – divided in between and within animal relationship.

		genomic relationship			pedigree based relationship		
		ref	val	between ref and val	ref	val	between ref and val
	N	2086	400	2486	2086	400	2486
relationship between animals	Min	-0.180	-0.146	-0.165	0.000	0.030	0.005
	Max	0.793	0.677	0.717	0.686	0.633	0.641
	Mean	0.0002	0.016	-0.003	0.093	0.120	0.096
	SD	0.055	0.071	0.049	0.052	0.068	0.044
relationship within animals	Min	0.824	0.878		1.000	1.027	
	Max	1.209	1.222		1.169	1.109	
	Mean	0.991	0.989		1.043	1.053	
	SD	0.054	0.056		0.018	0.012	

3.2 Methods

3.2.1 Variance components

The variance components were estimated based on a multi-trait animal model in ASReml (Gilmour et al. 2009). The model response variables included the deregressed breeding values of the traits NBA, UWP, PS and IWO with their corresponding weights. The weights are an analogue to the reliability of breeding values. The random effect animal accounts for the pedigree based relationship. Additive genetic (co)variances and residual (co)variances were estimated by ASReml and used to compute heritability, genetic and phenotypic correlations.

3.2.2 Statistical analysis

3.2.2.1 Genomic breeding values

The method GBLUP was used to calculate the genomic breeding values in ASReml (Gilmour et al. 2009). More background information regarding the statistical models used in GS can be found in section *2.3 Genomic Selection.* The results of this analysis was used as a reference to compare the other models. Therefore, it will be further referred to as the traditional GBLUP or default GBLUP.

The single-trait GBLUP model is:

$$y = 1_n \mu + Z g_1 + e \qquad (1.0)$$

y = vector of drEBVs

1_n = vector of 1s

μ = overall mean

Z = design matrix to link records to g_1

g_1 = vector of random effect based on the genomic relationship matrix (G) with assumed $N(0, I\sigma^2_{g_1})$

e = vector of residual errors with assumed $N(0, I\sigma^2_e)$

The genomic relationship matrix (G) is calculated, according to the method 1 of VanRaden (2008). The genotypes 0, 1 and 2 were standardised within Z_{ij} as $0 - 2p_j$, $1 - 2\,p_j$ and $2 - 2p_j$,

respectively for all animal i and markers j (Meuwissen et al. 2013). Accordingly, the standardization results in genotypes with mean zero. The method used to calculate G was:

$$G = Z * Z' \Big/ \sum_j 2 * p_j * (1 - p_j)$$

Z = matrix of standardized SNP genotypes Z_{ij} for animal i at marker$j$; with n rows (number of individuals) and m columns (number of SNPs)

Z' = transpose of Z

p_j = allele frequency of the minor allele at marker j

The division by $\sum_j 2p_j * (1 - p_j)$ gives more weight to alleles with higher minor allele frequency (MAF) (VanRaden 2008). The G-matrix is a $n\ x\ n$ matrix with between animals relationships on the off-diagonal elements and the within animal relationship on the diagonal. Furthermore, the division scales G to be analogous to the numerator relationship matrix A. For instance, can the inbreeding coefficient be obtained from the G-matrix by subtracting 1 of an individual's genomic relationship to itself ($G_{ii} - 1$).

The software used to calculate the G-matrix in this work was gghat version 3 from Meuwissen (2015).

3.2.2.2 Genomic breeding values with additional marker effect

Two different models were used to include the additional genomic information in the calculation of GEBVs in ASReml (Gilmour et al. 2009).

On one hand, the markers based on the QTLs from the database were used to build a genomic relationship matrix (QTL-matrix). The G-matrix based on all QTLs will further be called QTL-matrix and the G-matrix based on the reproduction QTLs will be called rQTL-matrix (Table 2). The matrix was built analogous to the G-Matrix discussed above. The model considered was:

$$y = 1_n\mu + Z_2 g_2 + Z_1 g_1 + e \qquad (2.1)$$

y = vector of drEBVs

1_n = vector of 1s for all n animals

μ = overall mean

Z_2 = design matrix to link records to g_2

g_2 = vector of random effect related to the genomic relationship matrix established from the marker of the QTL database (QTL-matrix or rQTL-matrix) with assumed $N(0, I\sigma_{g_2}^2)$

Z_1 = design matrix to link records to g_1

g_1 = vector of random effect related to the genomic relationship matrix (G) with assumed $N(0, I\sigma_{g_1}^2)$

e = vector of residual errors with assumed $N(0, I\sigma_e^2)$

On the other hand, with the fewer markers, it was possible to add them as single marker or haplotype, based on their phased information. For the smaller panel with one SNP, two different effect levels resulted, whereas the panel with two SNPs occurred in four different combinations. Further, it will be relate to the panels and models as *E.coli* panel/model and LS panel/model (Table 2). The new effect was added as diallelic fixed effects to the model:

$$y = 1_n\mu + (X_{pat} + X_{mat})h + Zg_1 + e \qquad (2.2)$$

y = vector of drEBVs

1_n = vector of 1s

μ = overall mean

Z = design matrix connecting records to the breeding value

X_{pat} = design matrix connection the animal record with the allelic effect inherited from the sire

X_{mat} = design matrix connection the animal record with the allelic effect inherited form the dam

h = vector of additive genetic effect for each marker or haplotype

g_1 = vector of additive genetic effect using the genomic relationship matrix (G) with assumed $N(0, I\sigma_{g_1}^2)$

e = vector of random residual errors with assumed $N(0, I\sigma_e^2)$

3.2.2.3 Evaluation

In order to calculate the present the correctness of the GEBVs the response variables of the 400 youngest animals were masked in the analysis. Further, the correlations between the estimated breeding values and the original deregressed breeding values of the 400 validation animals were calculated to assess the predictive ability of the models as:

$$\rho_{drEBV,\widehat{EBV}} = corr\left(drEBV, \widehat{EBV}\right)$$

Regarding the default GBLUP $\widehat{EBV}$ is simply the estimated effect for each animal i ($\hat{g}_{1i}$). Concerning the QTL and rQTL model $\widehat{EBV}$ has to be calculated as the sum of the effects of the QTL-matrix ($\hat{g}_{2i}$) and the G-matrix ($\hat{g}_{1i}$) for each animal i. The $\widehat{EBV}$ of the marker models were calculated as the sum of the two effects of the haplotype or marker ($\hat{h}_m$), depending on the effect level m, and the animal effect based on the G-matrix ($\hat{g}_{1i}$).

The prediction accuracies were obtained by the formula

$$accuracy = \frac{\rho_{drEBV,\widehat{EBV}}}{r^2_{drEBV}},$$

with $\rho_{drEBV,\widehat{EBV}}$, the correlation between the deregressed breeding value and the genomic estimated breeding value and divided by r^2_{drEBV}, the average reliability of the deregressed breeding values for the validation group (Table 3).

The significance of the QTL model and rQTL model was analysed based on their logarithmic likelihood. It was tested, if twice the absolute difference between the logarithmic likelihood of the QTL model and the logarithmic likelihood of the default GBLUP, was chi-square distributed.

To evaluate the effect of the new models on selection, the rank correlation was calculated. It is the correlation of the ranking based on the default GBLUP and the rank with each of the GBLUP models with additional marker effect.

3.2.3 Single marker analysis

To receive the significance and amount of genetic variance explained by the single markers used in the LS and *E.coli* model a SNP by SNP analysis was executed by Maren van Son. The same data and model 2.2 were used as introduced above. *E.coli* was examined regarding its importance on the trait PS and the LS markers regarding the trait NBA.

As this is not part of the project itself, the results are only indicated in the appendix (Table A5).

4 Results

4.1 Variance component

The genetic analysis of the traits indicated heritabilities of 0.21 for NBA, 0.16 for UWP, 0.17 for PS and 0.08 for IWO (Table 5). NBA and UWP show a strong positive genetic correlation (0.53), where an increase in litter size causes an increase in the number of underweighted piglets in a litter. All other traits show negative genetic correlations between each other, ranging from -0.003 to -0.53 (Table 5). The strongest negative correlation between PS and UWP implies that the more piglets are born underweighted the smaller is their chance to survive. Similar observations can be made for the phenotypic correlations (Table 5). The underlying additive genetic variances and residual variances are presented in Table 6. The additive genetic variance represents the part of phenotypic variation, which can be explained by the inheritance of genes.

*Table 5: Genetic parameters: **heritabilities (SE)** on the diagonal, genetic correlations (SE) under the diagonal and phenotypic correlations (SE) above the diagonal. For the traits litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO), based on all SLW animals available in the data set.*

	NBA		UWP		PS		IWO	
NBA	**0.21**	**(0.025)**	0.63	(0.026)	-0.14	(0.041)	-0.024	(0.038)
UWP	0.53	(0.039)	**0.16**	**(0.021)**	-0.47	(0.031)	-0.044	(0.037)
PS	-0.08	(0.054)	-0.53	(0.041)	**0.17**	**(0.021)**	-0.26	(0.035)
IWO	-0.003	(0.062)	-0.078	(0.065)	-0.10	(0.064)	**0.08**	**(0.011)**

Table 6: Additive genetic variance components σ_a^2 (SE) and residual variances σ_e^2 (SE).

	σ_a^2		σ_e^2	
Trait[1]				
NBA	0.98	(0.07)	3.79	(0.34)
UWP	5.34	(0.43)	27.37	(2.19)
PS	5.95	(0.48)	29.65	(2.38)
IWO	3.29	(0.34)	37.93	(2.17)

[1] *the traits are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO) based on all SLW animals available*

4.2 Using markers with large effects for the genomic predictions

The effect of including important markers additionally in GBLUP is evaluated. The marker panels were added in different extends. They ranged from including only one SNP to including 4,205 SNPs. Furthermore, they were added in different ways. The smaller panels were added as fixed effect and the big panels built G-matrices that were added as random effect.

4.2.1 Model comparison

The models are compared according to their likelihood values, which are stated in Table A6 in the appendix. The values named $2 * LogL$ are calculated as two times the absolute difference between the logarithmic likelihood of each QTL model and the logarithmic likelihood of the basic GBLUP model, presented in Table 7. The thresholds of 3.84 and 6.63 indicate significant improvement of models with a significance level of 0.05 and 0.01, respectively. Table 7 displays that the most models improve their likelihood, thus their goodness-of-fit with including a QTL-matrix, but only the model for NBA improves significantly with an additional rQTL-matrix.

Table 7: Model comparison between the QTL models and the default GBLUP model, regarding the chi-square distribution and the corresponding p-values for the traits litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO), including all SLW animals.

	NBA		UWP		PS		IWO	
	2*Logl	p-value	2*Logl	p-value	2*Logl	p-value	2*Logl	p-value
Model[1] Compared to GBLUP								
QTL	2.75	0.061	0	1	1.1	0.219	0	1
rQTL	7.28	0.004	0.32	0.601	0	1	0.54	0.414

[1] *the QTL models are the GBLUP model with additional QTL-matrix (QTL) and the GBLUP model with additional QTL-matrix associated with reproduction (rQTL)*

The effective amount of additive genetic variance that is explained in the models by the different genomic relationship matrices are shown in Table 8. The total genetic variance increased slightly for the trait NBA and IWO by including the rQTL-matrix. The QTL-matrix explains some of the additive genetic variance for the traits NBA and PS, whereas the rQTL-matrix also explains some of the variance for the trait UWP in addition to NBA and PS.

Table 8: The additive genetic variances explained by the different G-matrices, the total additive genetic variance and the residual variance for all models and traits used in the analysis of the breed SLW.

		Trait[2]			
Model[1]	Source of variance	NBA	UWP	PS	IWO
GBLUP	additive genetic variance	1.360	7.361	8.514	7.285
	Residual variance	1.345	11.407	17.980	6.299
QTL	G-matrix	1.253	7.361	8.121	7.285
	QTL-matrix	0.109	3.65E-06	0.395	1.80E-08
	total additive genetic variance	1.362	7.361	8.516	7.285
	Residual variance	1.344	11.410	17.990	6.299
rQTL	G-matrix	1.265	7.248	8.514	7.114
	rQTL-matrix	0.097	0.120	1.18E-06	0.179
	total additive genetic variance	1.362	7.368	8.514	7.293
	Residual variance	1.352	11.410	17.980	6.306
LS	additive genetic variance	1.360	7.370	8.518	7.297
	Residual variance	1.349	11.420	18.008	6.299
E.coli	additive genetic variance	1.361	7.366	8.412	7.288
	Residual variance	1.343	11.404	18.170	6.301

[1] the models used are the traditional GBLUP (GBLUP), GBLUP with additional QTL-matrix (QTL), GBLUP with additional reproduction QTL-matrix (rQTL), GBLUP with additional markers associated with litter size (LS) and GBLUP with the marker for E.coli resistance (E.coli)

[2] the traits are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO)

Table 9 presents the test statistics of the additional fixed effects, consisting of the *E.coli* marker and the LS markers. For the most traits, the additional marker effect does not have a significant influence. However, the *E.coli* marker has a significant impact on the proportion of nursed individuals that are weaned. The appendix includes a table showing the estimated effects of each effect level for the *E.coli* marker and the LS haplotype (Table A7). According to trait definitions, the number of piglets born alive (NBA) and the survival rate (PS) should be as high as possible,

whereas the number of piglets underweighted (UWP) and the interval weaning to oestrus (IWO) should be as short as possible. Regarding the *E.coli* marker, the frequency of the allele favouring PS has a frequency of 0.39. The same allele indicates improvements for NBA and UWP, but disadvantages for IWO (Table A7).

Table 9: ASReml output regarding the additional fixed effect in the single-trait GBLUP for all animals, based on markers associated with the number of piglets born alive (LS) and the E.coli marker (E.coli) for all SLW animals in the data set

	LS			*E.coli*		
	DF	F-value	P-value	DF	F-value	P-value
Trait[1]						
NBA	3	0.13	0.939	1	1.18	0.279
UWP	3	0.03	0.992	1	0.12	0.727
PS	3	0.22	0.881	1	8.03	0.005
IWO	3	0.35	0.786	1	0.05	0.812

[1] *the traits are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO)*

4.2.2 Comparison of the breeding values

The distributions of the breeding values predicted for the validation group in single-trait models are visualised in Figure 6. The distributions do not change much by including additional effects. The mean for the trait NBA is slightly lower by including the *E.coli* marker.

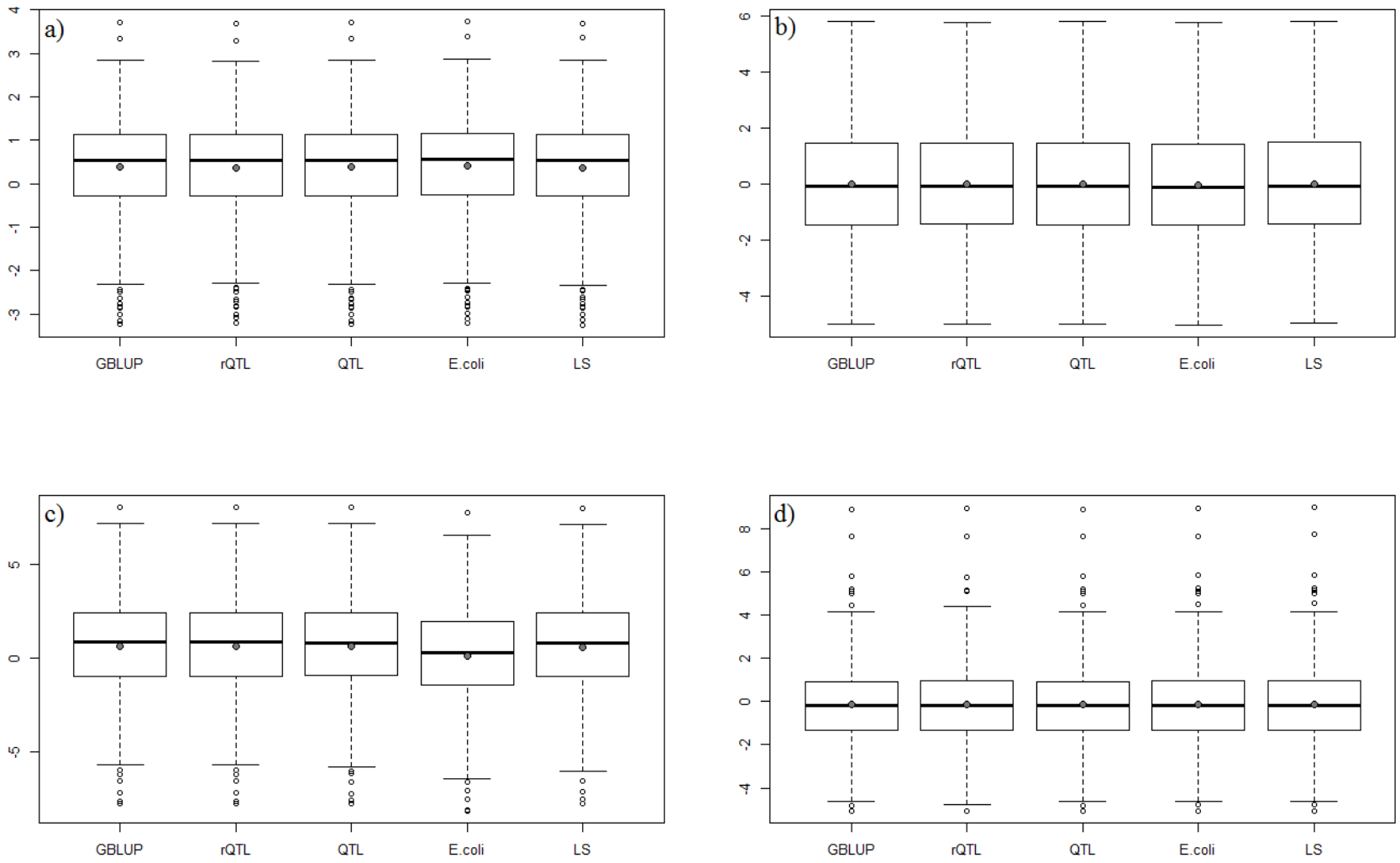


Figure 6: The boxplots show the distribution of the GEBVs, where the dots in the box indicate the mean. The models underlying the breeding value estimation are the default GBLUP (GBLUP), GBLUP with additional reproduction QTL-matrix (rQTL), GBLUP with additional QTL-matrix (QTL), the GBLUP with additional fixed effect for the E.coli marker (E.coli) and the GBLUP with additional fixed effect for the LS haplotype (LS). The subdivision is ordered as a) GEBVs for the number of animals born alive (NBA), b) GEBVs for the trait proportion of underweighted piglets (UWP), c) GEBVs for the trait survival rate (PS), and d) GEBVs for the trait interval weaning to oestrus (IWO).

4.2.3 Prediction ability

The models ability to predict breeding values accurate is shown in Table 10 and Table A8 in the appendix. The correlations between the GEBVs of the 400 masked animals and their corresponding deregressed breeding values are used to compare the prediction ability for the new models. The models with additional QTL-matrix improve the correlation slightly for NBA compared to the default GBLUP model (Table 10). Adding the reproduction associated relationship matrix improves the predictive ability to maximal 0.0013 (for NBA) in the current data. However, the models including the LS and the *E.coli* markers show similar correlations between the GEBVs and the deregressed EBVs as the traditional GBLUP. The prediction abilities for the GBLUP models indicate accuracies over 100% (Table A8).

Table 10: Correlations between the genomic estimated breeding values (GEBVs) and the obtained deregressed breeding values for the 400 masked animals for each statistical model (the default GBLUP, the GBLUP with additional QTL-matrix (QTL), the GBLUP with additional rQTL-matrix (rQTL), the GBLUP with additional litter size associated haplotypes (LS) and the model with additional fixed effect of the E.coli marker (E.coli)).

	GBLUP	QTL	rQTL	LS	*E.coli*
Trait[1]					
NBA	0.8429	0.8434	0.8442	0.8426	0.8432
UWP	0.7856	0.7856	0.7857	0.7855	0.7857
PS	0.7859	0.7853	0.7859	0.7847	0.7849
IWO	0.6841	0.6837	0.6846	0.6836	0.6841
mean[2]	0.7746	0.7745	0.7751	0.7741	0.7745

[1] *the traits are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO)*

[2] *average performance of the model*

4.2.4 Impact on the ranking

To assess the reranking according to the GEBVs, the correlations between the ranking in the default GBLUP model and the models with additional genetic information were computed for the 400 masked animals (Table 11). The correlations are all above 0.99, where there is almost no difference.

Table 11: Rank correlations calculated as the correlation between the ranks according to the default GBLUP and the new introduced models regarding the reproduction traits for the 400 validation animals. The models are the GBLUP with additional QTL-matrix (QTL), GBLUP with additional reproduction QTL-matrix (rQTL), GBLUP with fixed effect for markers associated with litter size (LS) and GBLUP with additional fixed effect for the E.coli marker (E.coli)

	QTL	rQTL	LS	*E.coli*
Trait[1]				
NBA	1	0.9988	0.9999	1
UWP	1	0.9999	0.9999	1
PS	0.9996	1	0.9998	0.9991
IWO	0.9999	0.9994	0.9996	1

[1] *the traits are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO)*

5 Discussion

Genetic parameter

Genetic parameters are used to describe genetic qualities of traits and are needed to predict breeding values. The genetic parameters from Table 5 and 6 indicate that there is sufficient additive genetic variance to improve the traits NBA, UWP, PS and IWO. The heritabilities computed are generally higher than the values obtained by SUISAG (Hofer 2016a), where NBA has a heritability of 0.12, UWP 0.09, PS 0.06 and IWO 0.12. Furthermore, the additive genetic variances and the residual variances are both lower than usual (Hofer 2016a). An explanation for these differences is the restricted sample size, as in this analysis only animals with phenotypic observations as well as genomic information are included. This is a weighted analysis, using deregressed proofs corrected for the parent average and the shrinking within BLUP, thus it is possible that measurements of the descendants have an impact on the outcome. Even though Garrick et al. (2009) claims the impact to be negligible for genomic predictions, it cannot be assumed that it does not affect the estimation of the variance components and heritabilities in this study.

In general, the heritabilities are either slightly above or within the area of the heritabilities found in literature (Chen et al. 2003; Hanenberg et al. 2001; Lee et al. 2015; Nguyen et al. 2003; Putz et al. 2015; Serenius et al. 2008). The genetic correlations obtained in the study showed similarities and contradictions to the ones in literature. Negative genetic correlations between NBA and average birth weight had been published by Nguyen et al. (2003) and Putz et al. (2015), what supports the positive correlation found between NBA and UWP. Because UWP represents a proportion, the correlation becomes positive, when the birth weights go down. The results indicate that an increase in the number of piglets born alive leads to a decrease in the average birth weight and accordingly to an increase in the proportion of piglets born underweighted. Nguyen et al. (2003) supports the correlation between UWP and PS, as they show positive correlations (0.27) between birth weight and the number of piglets weaned. The conclusion drawn is that piglets with higher birth weight and a lower number of piglets born underweighted have a greater chance to survive.

Some genetic correlations calculated in this research project revealed opposing results compared to the literature. Contradictory to the small and negative correlation between NBA and PS in this

data, Putz et al. (2015) showed a high positive correlation (0.74) in their Large White population. Another contradictory element is the correlation between PS and IWO, that Serenius et al. (2008) found to be slightly positive instead. Furthermore, the positive correlation between NBA and IWO, described by Serenius et al. (2008) and by Wilson and Dewey (1993) could not be confirmed in this study. All these contradictions between the genetic parameters found in this research and the values from the literature occur because they are analysed on different breeds, different sizes of data sets, different traits and methods to measure the phenotypes. Nevertheless, the genetic correlations indicate, that by selecting for one trait the other traits will be affected. For example, by selecting only for a higher number of animals born alive the proportion of piglets underweighted will raise and consequently, the survival rate will decrease. By including all traits in the breading goal and selecting them simultaneously, the outcome can be balanced and lead to a satisfying improvement for all traits.

Additional genomic relationship matrix

Generally, the additional genomic relationship matrices do improve the likelihood of the normal GBLUP model. However, for the data used it is not significant in the most cases. Solely for the trait NBA, the model showed significant improvement by adding a reproduction associated QTL-matrix additionally to the G-matrix, as was shown in Table 7. Even though significant improvements for the QTL models were shown, the prediction accuracies do not respond in noteworthy changes.

In order to discuss the results it is vital to understand, how the two genomic matrices in the model relate to each other. Table 8 shows that the variance explained by the G-matrix is higher than the variance explained by a QTL-matrix. It proves to be more beneficial to include all genes rather than including few important genes. Even though, they are overlapping, the additional weight put on the QTL-markers does not result in outstanding enhancements. An explanation is, that the G-matrix accounts for enough of the QTL-markers' variance and covariance, that the effect of the QTL-matrix decreases. Furthermore, the number of effects that have to be estimated doubles by including a second relationship matrix. With more effects estimated more errors are included and the accuracy of a model decreases. In this study, a noteworthy improvement could only be

detected for the likelihood of the model with the rQTL-matrix and trait NBA, but not for the prediction accuracy.

In a similar study, Brøndum et al. (2015) compared the inclusion of QTLs as an additional variance component in a GBLUP and Bayesian model. They showed an increase up to 2.2 percentage points of the prediction accuracy for breeding values with the additional QTL-matrix included in the GBLUP model (Brøndum et al. 2015). More precisely, they built a G-matrix based on a 54k SNP-panel and a QTL-matrix based on 1,623 SNPs that were not included in the G-matrix. It was compared to a GBLUP model, including a G-matrix with all markers. Therefore, they can observe an increase, whereas in the current study, the QTLs are included twice in the model and the effect of the QTL-matrix is reduced.

In general, the rQTL-matrix works better than the QTL-matrix. The only difference occurs for the trait PS, where amongst other possible reasons the QTL-matrix includes the *E.coli* marker, but it is not included in the rQTL-matrix. Brøndum et al. (2015) made similar observations by comparing nonspecific QTL-panels and specific QTL-panels regarding breed and traits. The limitation for breed and trait specific markers resulted in an improvement of the prediction accuracy for some breeds. Adding non-specific QTL-markers includes noise and dilutes the QTL-relationship regarding the specific trait (Brøndum et al. 2015).

Concerning the differences in outcomes between traits, the trait association of the QTLs is very uneven distributed across the panels. The QTL database includes many QTLs for the number of piglets born alive, fewer for birth weight, hardly any for the number of piglets weaned and no QTLs for the interval weaning to oestrus. Therefore, marker selection in this part of the study is not very trait specific. One could have been more restrictive regarding the trustworthiness of the QTLs. For instance, including only regions with several known QTLs affirmed by independent studies or checking if the QTLs from the database are relevant for the traits of interest and the population used.

Moreover, the selection of the traits themselves affects the outcome. Brøndum et al. (2015) showed that predictions improved more for production than for reproduction traits. They explain it with the complexity of reproduction. Reproduction traits are influenced by the expression of more genes than production traits and have lower heritabilities. As in the current study only

reproduction traits are concerned, it would be interesting to see if the results improve for production traits with high heritabilities.

Additional markers as fixed effect

The first of the two models with additional fixed effect in GBLUP included two markers that are associated with the number of piglets born alive. The effect of the two-SNP haplotype showed no significance and did not improve the default GBLUP model, shown in Table 9. Its significant impact on the trait NBA found in the paper Bergfelder-Druing et al. (2015) could not be confirmed with the data used. Boichard et al. (2012) analysed, how the additional random effect of haplotypes improves a BLUP estimation. They achieved a slightly better accuracy with haplotypes associated to QTLs in BLUP, than with the normal GBLUP. Based on this results it should be discussed if the haplotypes and marker should have been added as random effects.

For this study, it was decided to treat them as fixed effects, considering the specific interest in the effects of each haplotype combination and each allele for the *E.coli* resistance. Furthermore, the SNPs are not chosen randomly. It is known that these specific markers are of importance and should therefore be fixed effects. When more markers should be implemented, the inclusion of haplotypes as random effects might be more interesting. In that way, it is possible to account for their variance and covariance.

The second model built, included an additional fixed effect in form of a single marker and haplotypes in GBLUP. The *E.coli* marker shows a significant effect in the single-trait model for the proportion of piglets born that are weaned (Table 9). Further, the prediction accuracy did not increase and the rank correlation shows no considerable changes. Additionally, the association analysis of *E.coli* with PS showed 2.8% of the genetic variance explained by the selected marker (Table A5). Contradictory to the ASReml output is, that the allele found to be the resistant version, does not correspond to the allele found by Neuenschwander et al. (2013). Anyhow, this should not affect the statistical outcome of the study. *E.coli* does not have a direct phenotype, as it is too expensive to test each herd for their *E.coli* occurrence. Nonetheless, *E.coli* resistance can be associated with PS, and thus PS could be used as indirect phenotype. As PS is included in the breeding programme, selection for *E.coli* resistant pigs already occurs unintended and the marker frequency is supposed to increase for this particular SNP.

Lopes et al. (In Press) added single SNPs that were highly associated with the trait number of teats in a single-trait GBLUP. The genetic variance explained by the main marker was 3.30% – 6.13%, depending on the breed. They achieved an increase in the prediction accuracy ranging from 0.003 to 0.043 in absolute values varying between breeds (Lopes et al. In Press). In the present study, the accuracy could not be increased with including the *E.coli* marker in the breeding value estimation of PS. The most possible explanation for the inconsistent results is that the *E.coli* marker does not have a strong enough effect to improve the prediction accuracy. Lopes et al. (In Press) defined the threshold to divide between important and unimportant markers for MA-GS as a minimal variance explained by the single marker of >1%. If a SNP explains less than 1% of the additive genetic variance, its inclusion in MA-GBLUP will not improve the model. This has happened with the LS markers in the data set, as it was not significant in the association study with NBA (results shown in appendix Table A5), hence did not affect the outcome of the models analysed.

Generally, the correlations between the predicted and the deregressed breeding values are relatively high (Table 10). It has been proven that the pedigree structure can affect the prediction accuracy of breeding values. A high relationship between the reference and the validation group favours the reliability of genomic predictions (Habier et al. 2007; Wu et al. 2015). The pedigree of SLW shows an average estimated relationship between the reference and the validation group of 0.096 (Table 4). That lays between cousins (0.125) and second cousins (0.031). Therefore, the predictive ability is quite high for SLW. Analysing other populations will result in different extents of the prediction abilities. Furthermore, the values in Table A8 show more than 100% accuracy of the predictions of the genomic models. This occurs, if the reliabilities are underestimated. Another explanation could be that the residuals are not independent and thus predictable.

The selected SNP-panels added to the models were not excluded from the panel used to build the G-matrix and are therefore twice included in the estimation. Consequently, the models could show problems with double counting or overfitting. In a trial, Lopes et al. (In Press) excluded the markers they included additionally to the model from the panel to build the G-matrix. The results

did not show noteworthy changes. However, a few double included single SNPs do not affect the model, but a whole panel of SNPs might complicate the calculations in a larger data set. Another open question is how the models will behave in a multi-trait GBLUP. Multi-trait models are interesting regarding correlated traits. The inclusion of major QTLs could have beneficial effects for all traits. Considering, that the multivariate model is computationally demanding, an additional QTL-matrix will double the resources needed for the computation.

Other approaches

Many different methods are being and have been tested to improve the prediction ability of GBLUP. One method analysed by Su et al. (2014) is to weight single markers or group of markers within the G-matrix used in GBLUP. They weighted selected markers with different variances obtained from a Bayesian model and GWAS. The results were up to 2 percent point improved reliabilities compared to the unweighted GBLUP (Su et al. 2014). Similar increases were achieved by weighting the markers with $-\log_{10}(p_j)$, with p_j representing the p-value obtained by an association study for SNP j (de Los Campos et al. 2013). Another approach proposed by Zhang et al. (2010) was to include a trait-specific relationship matrix in GBLUP. Therefore, the usual G-matrix is replace by a matrix that is weighted depending on their contribution explaining the variance of a trait. The authors showed improved prediction accuracies compared to the common mGBLUP and GBLUP models, but not compared to BayesB (Zhang et al. 2010).

Practical aspects of MA-GBLUP

Adding a major gene effect to the genomic breeding value estimation should be well planned. Firstly, the important SNPs for the population of interest have to be known. The possible improvement depend mainly on the trait itself. Concerning a trait that is easy to measure at an early stage, the improvement would not be immense, but for traits that are hard to measure or can only be measured on relatives, the improvements could be significant. Investing resources to discover markers that explain sufficient genetic variance is fundamental. Furthermore, the markers need to be available. If this is settled, the SNPs can simply be added to the marker file for the imputation. Regarding the implementation of single SNPs, the animals with missing

genomic information will need estimates for their genomic performance. Thus, the breeding values will be the sum of the marker effect and the animal effect based on the G-matrix.

When many SNPs, are to implement, as done with the QTL-markers from the database, two genomic relationship matrices are built and added to the breeding value estimation. Consequently, the computational demands will increase, since the number of calculations will double. The final breeding values are the sum of the effects for each animal in the respective G-matrix.

Association studies should be updated from time to time, considering the changes in allele frequencies with the alternation of generations. As soon as a marker is fixed in the population, it will not show any influence on the trait anymore, but other positions in the genome will become more important. Therefore, the markers of interest change over time.

6 Conclusion

Including major genes in GBLUP for reproduction traits in swine does improve the likelihood and thus the goodness-of-fit of the prediction models.

Adding non-specific QLTs did not improve prediction accuracy in the current data.

Giving QTL-markers more weight, by adding them to the GBLUP model, did not improve the prediction accuracy of GEBVs.

Even though, this work could not expose any significant improvement of the prediction accuracy, the goodness-of-fit was improved and a larger, more powerful data set might convert the improved goodness-of-fit to a better accuracy of prediction.

7 References

Bergfelder-Druing, S., Grosse-Brinkhaus, C., Lind, B., Erbe, M., Schellander, K., Simianer, H. & Tholen, E. (2015). A Genome-Wide Association Study in Large White and Landrace Pig Populations for Number Piglets Born Alive. *Plos One*, 10 (3): 1-19. e011746.

Boichard, D., Guillaume, F., Baur, A., Croiseau, P., Rossignol, M. N., Boscher, M. Y., Druet, T., Genestout, L., Colleau, J. J., Journaux, L., et al. (2012). Genomic selection in French dairy cattle. *Animal Production Science*, 52 (2-3): 115-120.

Brøndum, R. F., Su, G., Janss, L., Sahana, G., Guldbrandtsen, B., Boichard, D. & Lund, M. S. (2015). Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J Dairy Sci*, 98 (6): 4107-4116.

Browning, B. L. & Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*, 84 (2): 210-223.

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

Chen, P., Baas, T. J., Mabry, J. W., Koehler, K. J. & Dekkers, J. C. (2003). Genetic parameters and trends for litter traits in U.S. Yorkshire, Duroc, Hampshire, and Landrace pigs. *J Anim Sci*, 81 (1): 46-53.

de Los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C. & Sorensen, D. (2013). Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet*, 9 (7): 1-15. e1003608.

dos Santos, J. P., Vasconcellos, R. C., Pires, L. P., Balestre, M. & Von Pinho, R. G. (2016). Inclusion of Dominance Effects in the Multivariate GBLUP Model. PLoS One, 11 (4): 1-21. e0152045.

Fairbrother, J. M., Nadeau, E. & Gyles, C. L. (2005). Escherichia coli in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. *Anim Health Res Rev*, 6 (1): 17-39.

Falconer, D. S. & Mackay, F. C. (1996). *Introduction to quantitative genetics*. 4rd ed. Burnt Mill, Harlow, Essex, England.

FAO (2016a). Biannual Report on Global Food Markets. *Food Outlook*, I5703. 1-139 pp.

FAO (2016b). *Pigs and Animal Production*. http://www.fao.org/Ag/againfo/themes/en/pigs/production.html (accessed: 25.11.16).

Fredholm, M. (2008). *Genetic improvement of porcine health*. International Pig Veterniary Society Congress, Durban - South Africa, pp. 1-4: Genetics and Bioinformatics, IBHV, The Royal Veterinary and Agricultural University, Denmark.

Frey, M. (1999). *Zuchtwertschätzung für Reproduktionsmerkmale beim Schwein*. Zürich: Swiss Federal Institute of Technology in Zurich, Departement of Agricultural Sciences. 123 pp.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V. K., Weiler, J. E., O'Brien, P. J. & MacLennan, D. H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science*, 253 (5018): 448-51.

Garrick, D. J., Taylor, J. F. & Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol*, 41: 55.

Gilmour, A. R., Gogel, B., Cullis, B. & Thompson, R. (2009). *ASReml user guide release 3.0*. VSN International Ltd, Hemel Hempstead, UK.

Goddard, M. E., Hayes, B. J. & Meuwissen, T. H. (2010). Genomic selection in livestock populations. *Genet Res (Camb)*, 92 (5-6): 413-421.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., et al. (2002). Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, 12 (2): 222-231.

Guerin, G., Duval-Iflah, Y., Bonneau, M., Bertaud, M., Guillaume, P. & Ollivier, L. (1993). Evidence for linkage between K88ab, K88ac intestinal receptors to Escherichia coli and transferrin loci in pigs. *Anim Genet*, 24 (5): 393-396.

Guo, G., Lund, M. S., Zhang, Y. & Su, G. (2010). Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. *Journal of Animal Breeding and Genetics*, 127 (6): 423-432.

Habier, D., Fernando, R. L. & Dekkers, J. C. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177 (4): 2389-2397.

Hanenberg, E. H. A. T., Knol, E. F. & Merks, J. W. M. (2001). Estimates of genetic parameters for reproduction traits at different parities in Dutch Landrace pigs. *Livestock Production Science*, 69 (2): 179-186.

Hellbrugge, B., Tolle, K. H., Bennewitz, J., Henze, C., Presuhn, U. & Krieter, J. (2008). Genetic aspects regarding piglet losses and the maternal behaviour of sows. Part 1. Genetic analysis of piglet mortality and fertility traits in pigs. *Animal*, 2 (9): 1273-1280.

Henckel, P., Jorgensen, P. F. & Jensen, P. (1992). Glycogen-Content, Buffering Capacity and Resting Ph in Live Muscles of Pigs of Different Halothane Genotypes (a Pilot Project). *Meat Science*, 32 (2): 131-138.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31 (2): 423-447.

Hidalgo, A. M., Bastiaansen, J. W., Lopes, M. S., Veroneze, R., Groenen, M. A. & de Koning, D. J. (2015). Accuracy of genomic prediction using deregressed breeding values estimated from purebred and crossbred offspring phenotypes in pigs. *J Anim Sci*, 93 (7): 3313-3321.

Hofer, A. (2016a). *Genetische Parameter ZWS*. SUISAG (ed.).

Hofer, A. (2016b). Zuchtfortschritt in der Aufzuchtleistung. *Suisseporcs Information*, 6/2016: 16-18.

Hu, Z. L., Dracheva, S., Jang, W., Maglott, D., Bastiaansen, J., Rothschild, M. F. & Reecy, J. M. (2005). A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm Genome*, 16 (10): 792-800.

Hu, Z. L., Park, C. A. & Reecy, J. M. (2016). Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res*, 44 (D1): D827-833.

Jacobsen, M., Kracht, S. S., Esteso, G., Cirera, S., Edfors, I., Archibald, A. L., Bendixen, C., Andersson, L., Fredholm, M. & Jorgensen, C. B. (2010). Refined candidate region specified by haplotype sharing for Escherichia coli F4ab/F4ac susceptibility alleles in pigs. *Anim Genet*, 41 (1): 21-25.

Jermiah, L. E., Gibson, J. P., Gibson, L. L., Ball, R. O., Aker, C. & Fortin, A. (1999). The influence of breed, gender, and PSS (Halothane) genotype on meat quality, cooking loss, and palatability of pork. *Food Research International*, 32 (1): 59-71.

Jia, Y. & Jannink, J. L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, 192 (4): 1513-1522.

Jonas, E. & de Koning, D. J. (2015). Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. *Front Genet*, 6: 49. doi: 10.3389/fgene.2015.00049.

Jorgensen, C. B., Cirera, S., Anderson, S. I., Archibald, A. L., Raudsepp, T., Chowdhary, B., Edfors-Lilja, I., Andersson, L. & Fredholm, M. (2003). Linkage and comparative mapping of the locus controlling susceptibility towards E. COLI F4ab/ac diarrhoea in pigs. *Cytogenet Genome Res*, 102 (1-4): 157-162.

Lande, R. & Thompson, R. (1990). Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genetics*, 124 (3): 743-756.

Lee, J. H., Song, K. D., Lee, H. K., Cho, K. H., Park, H. C. & Park, K. D. (2015). Genetic Parameters of Reproductive and Meat Quality Traits in Korean Berkshire Pigs. *Asian-Australas J Anim Sci*, 28 (10): 1388-1393.

Legarra, A., Robert-Granie, C., Croiseau, P., Guillaume, F. & Fritz, S. (2011). Improved Lasso for genomic selection. *Genet Res (Camb)*, 93 (1): 77-87.

Lillehammer, M., Meuwissen, T. H. & Sonesson, A. K. (2011). Genomic selection for maternal traits in pigs. *J Anim Sci*, 89 (12): 3908-3916.

Lillehammer, M., Meuwissen, T. H. & Sonesson, A. K. (2013). Genomic selection for two traits in a maternal pig breeding scheme. *J Anim Sci*, 91 (7): 3079-3087.

Lopes, M. S., Bovenhuis, H., van Son, M., Nordbo, Ø., Grindflek, E., Knol, E. F. & Bastiaansen, J. (In Press). Using markers with large effect in genetic predictions. *J Anim Sci*: 32.

Luther, L. (2015). Resistenzzucht schreitet weiter voran. *Suisseporcs Information*, 9/2015: 16-17.

Meijerink, E., Neuenschwander, S., Fries, R., Dinter, A., Bertschinger, H. U., Stranzinger, G. & Vogeli, P. (2000). A DNA polymorphism influencing alpha(1,2)fucosyltransferase activity of the pig FUT1 enzyme determines susceptibility of small intestinal epithelium to Escherichia coli F18 adhesion. *Immunogenetics*, 52 (1-2): 129-136.

Meuwissen, T., Hayes, B. & Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annu Rev Anim Biosci*, 1: 221-37.

Meuwissen, T. (2015). *gghatvr3*. Gghat. http://wiki.nmbu.org/index.php/Gghat_version_3: NMBU Bioinformatics.

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

Meuwissen, T. H. E. & Goddard, M. E. (1996). The use of marker haplotypes in animal breeding schemes. *Genetics Selection Evolution*, 28 (2): 161-176.

Nagy, B. & Fekete, P. Z. (2005). Enterotoxigenic Escherichia coli in veterinary medicine. *Int J Med Microbiol*, 295 (6-7): 443-454.

Neuenschwander, S., Vögeli, P., Bertschinger, H. U., Rampoldi, A., Bratus, A., Stüssi, M., Bürgi, E. & Sidler, X. (2013). Genetische Resistenz gegen ColiF4 Bakterien: DNA-Marker vorhanden. *Suisseporcs Information*, 2/2013: 21-23.

Nguyen, N. H., McPhee, C. P. & Wade, C. M. (2003). Genetic parameters for reproduction traits in sows of lines selected for growth on restricted feeding. *Proc. Adv. Anim. Breed.Gen.*, 15: 245-248.

Nielsen, B. & Johannsen, M. (2004). *Sammenhæng mellem genotype for E.coli 0149-F4ab/ac og egenskaberne i avlsmålet*. Videncenter for Svineproduktion. Electronic article. Accessed under: http://vsp.lf.dk/Publikationer/Kilder/lu_medd/2004/678.aspx?full=1

Nielsen, B. & Johannsen, M. N. (2005). F4-resistente grise har højere avlsindeks. *DANAVL MAGASINET*, February 2016: 13.

Nielsen, B. & Svensmark, B. (2010). Effekten af F4-avl. *DANAVL MAGASINET*, Maj 2010: 1-2.

Nordbø, Ø., GJerlaug-Enger, E., Aasmundstad, T. & Grindflek, E. (2014). *Genomic Relationships Give Improved Prediction Ability for a Selection of Traits in Norsvin Landrace and Duroc*. World Congress on Genetics Applied to Livestock Production, Vancouver, Canada, p. 3: Norsvin & Geno SA, Hamar, Norway.

Ostersen, T., Christensen, O. F., Henryon, M., Nielsen, B., Su, G. S. & Madsen, P. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution*, 43: 38.

PigQTLdb (2016). *PigQTLdb*. http://www.animalgenome.org/cgi-bin/QTLdb/SS/index (accessed: 7.9.2016).

Pimentel, E. C. G., Wensch-Dorendorf, M., Konig, S. & Swalve, H. H. (2013). Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture. *Genetics Selection Evolution*, 45: 12.

Pimentel, E. C. G., Edel, C., Emmerling, R. & Gotz, K. U. (2015). How imputation errors bias genomic predictions. *Journal of Dairy Science*, 98 (6): 4131-4138.

Putz, A. M., Tiezzi, F., Maltecca, C., Gray, K. A. & Knauer, M. T. (2015). Variance component estimates for alternative litter size traits in swine. *Journal of Animal Science*, 93 (11): 5153-5163.

Rampoldi, A., Jacobsen, M. J., Bertschinger, H. U., Joller, D., Burgi, E., Vogeli, P., Andersson, L., Archibald, A. L., Fredholm, M., Jorgensen, C. B., et al. (2011). The receptor locus for Escherichia coli F4ab/F4ac in the pig maps distal to the MUC4-LMLN region. *Mamm Genome*, 22 (1-2): 122-19.

Sargolzaei, M., Chesnais, J. P. & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15: 478.

Schroyen, M., Stinckens, A., Verhelst, R., Niewold, T. & Buys, N. (2012). The search for the gene mutations underlying enterotoxigenic Escherichia coli F4ab/ac susceptibility in pigs: a review. *Vet Res*, 43: 70.

Serenius, T., Stalder, K. J. & Fernando, R. L. (2008). Genetic associations of sow longevity with age at first farrowing, number of piglets weaned, and wean to insemination interval in the Finnish Landrace swine population. *Journal of Animal Science*, 86 (12): 3324-3329.

Su, G., Christensen, O. F., Janss, L. & Lund, M. S. (2014). Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *Journal of Dairy Science*, 97 (10): 6547-6559.

SUISAG (2015a). *Schweizer Edelschwein ES - Für die Jungsauenerzeugung*. http://www.suisag.ch/Zucht/Rassen/Edelschwein (accessed: 22.8.16).

SUISAG (2015b). *Schweizer Zuchtprogramm - Effizienteste Produktion von Qualitäts-Mastschweinen*. http://www.suisag.ch/Zucht/Zuchtprogramm/tabid/80/Default.aspx (accessed: 22.8.2016).

SUISAG (2015c). Zahlen und Projekte. Sempach, LU, Switzerland: Aktionegesellschaft für Dienstleistungen in der Schweineproduktion. 36 pp.

SUISAG (2016). *Gewichtung der Merkmale in Teilindices und Gesamtzuchtwert im Jahr 2016 definitiv*. http://www.suisag.ch/Zucht/Zuchtprogramm/Zuchtziele/tabid/317/Default.aspx. p. 1. (accessed: 22.8.2016).

Van Laere, A. S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A. L., Haley, C. S., Buys, N., Tally, M., et al. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, 425 (6960): 832-836.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J Dairy Sci*, 91 (11): 4414-4423.

Velasco, S. C. H. (2011). *Genetics of Litter Siz and Prenatal Survival in pigs*: University of Ediburgh. 1-326 pp.

Wilson, M. R. & Dewey, C. E. (1993). The associations between weaning-to-estrus interval and sow efficiency. *Swine Health and Production*, 1 (4): 10-15.

Wu, X., Lund, M. S., Sun, D., Zhang, Q. & Su, G. (2015). Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. *Journal of Animal Breeding and Genetics*, 132 (5): 366-375.

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D. J. & Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One*, 5 (9). e12648.

8 Appendix

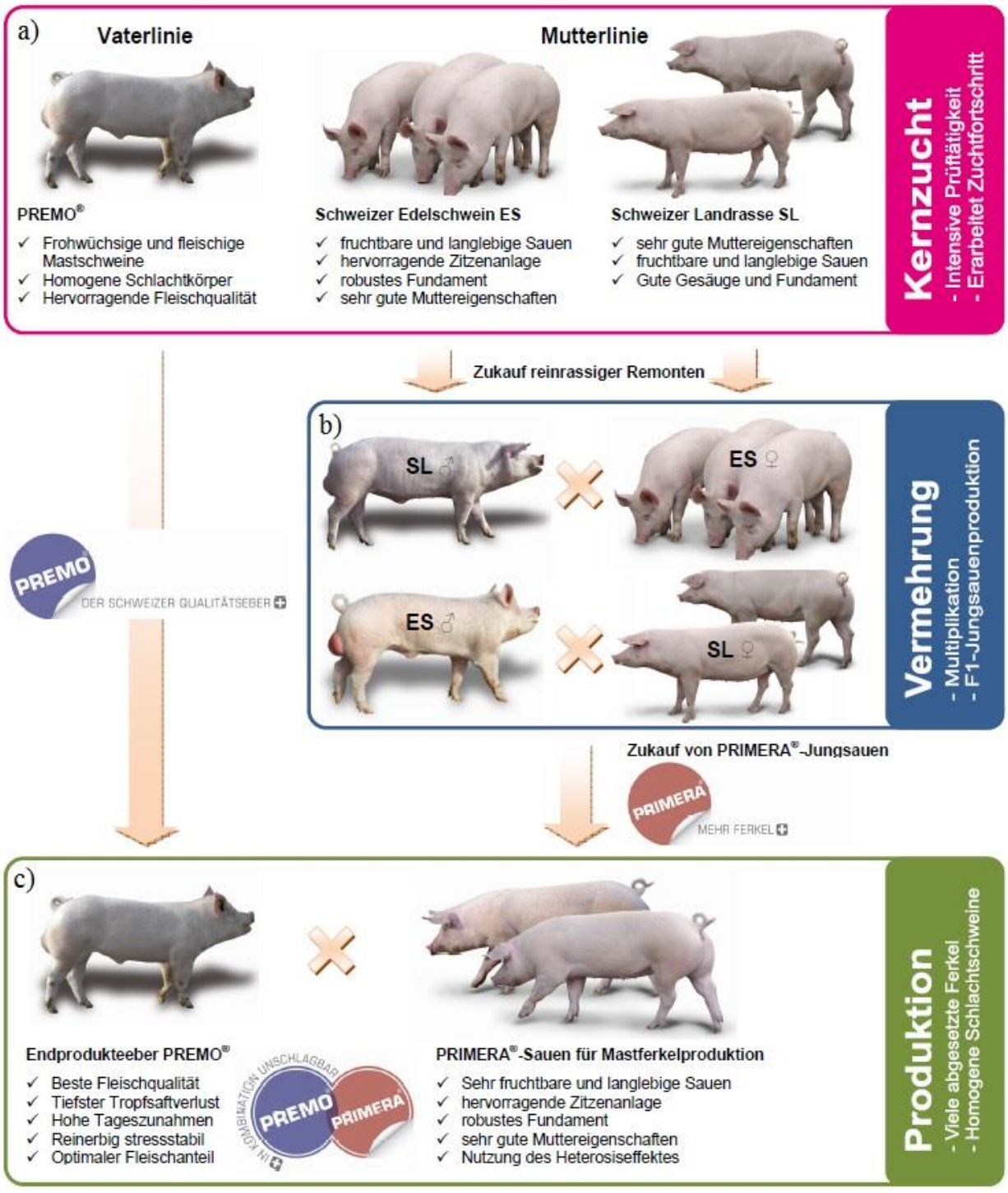


Figure A1: Breeding programme of SUISAG (http://www.suisag.ch/Zucht/Zuchtprogramm/tabid/80/Default.aspx). Displayed is the structure based on the three levels of breeding farms a) the nucleus herds, b) the multiplier herds and c) the production herds.

Table A1: Production traits, including each group with its single-traits and their relative weights for the SLW dam line and sire line.

		relative weight	
group	trait	dam line	sire line
weight gain	daily weight gain measured in mast	9	9
	daily weight gain measured in the field	11	6
	daily weight gain measured in the slaughterhouse	7	14
		27	29
meat quantity	loin eye area	3	8
	proportion lean meat	7	8
		10	16
meat quality	proportion intramuscular fat	19	13
	pH 1 hour post mortem	9	7
	pigment	4	3
	proportion drip loss	9	14
		41	37
feed efficiency	feed conversion	21	18
		21	18

Table A2: Conformation traits, including each group with its single-traits and their relative weights for the SLW dam line and sire line.

		relative weight	
group	trait	dam line	sire line
type	regularity of loin	7	10
	carcass length	5	2
		12	12
fundament	X-O rear leg	21	23
	side view angle rear leg	6	6
	angle pastern rear leg	2	0
	size of inner claws rear leg	18	24
	side view angle foreleg	3	3
	number of spots with liquid at joints	5	6
	gait	7	8
		62	70
teats	number of teats left	4	0
	number of teats right	4	0
	number of inverted teats	15	8
	number of intermediate teats	2	0
		25	8

Table A3: Statistical reference numbers (minimum – Min, maximum – Max, average – Mean and standard deviation – SD) of the phenotypic data set for all animals of the SLW. The columns are the estimated breeding values (EBV), the reliabilities of the estimated breeding values (r^2), the deregressed breeding values (drEBV), the reliabilities of the deregressed breeding values (r^2 drEBV) and the weight of the deregressed breeding values (weight)

Trait[1]		EBV	r^2	drEBV	r^2 drEBV	weight
NBA	Min	-1.42	0.38	-2.83	0.20	1.82
	Max	4.59	0.99	6.72	0.99	67.39
	Mean	1.99	0.67	2.06	0.57	11.90
	SD	0.96	0.11	1.36	0.16	11.97
UWP	Min	-4.36	0.30	-12.81	0.20	2.31
	Max	7.36	0.98	30.05	0.98	82.85
	Mean	1.25	0.64	1.28	0.54	13.43
	SD	1.64	0.11	2.97	0.15	13.35
PS	Min	-6.06	0.32	-11.38	0.18	3.27
	Max	8.07	0.99	12.22	0.99	132.06
	Mean	1.99	0.61	2.23	0.49	18.86
	SD	1.87	0.12	3.24	0.17	21.64
IWO	Min	-5.76	0.27	-14.87	0.13	1.01
	Max	5.54	0.98	16.54	0.98	58.92
	Mean	-2.68	0.54	-2.93	0.39	6.55
	SD	1.13	0.14	2.85	0.20	8.87

[1] *the traits are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO)*

Table A4: Number of markers selected for the QTL-matrix (QTL) and the reproduction associated QTL-matrix (rQTL) for each chromosome based on the pig QTL database.

	Number of SNPs	
chromosome	QTL	rQTL
1	344	112
2	350	84
3	164	55
4	343	68
5	170	41
6	346	56
7	362	97
8	328	89
9	247	63
10	123	38
11	90	22
12	211	49
13	219	60
14	277	61
15	216	85
16	192	40
17	123	55
18	100	28

Table A5: Description of the selected markers for the GBLUP models with additional fixed effect, including the chromosome and its allocated position, the minor allele frequency (MAF), the significance (p-value) in association with the trait, the allele substitution effect (a) and the genetic variance explained (var)

SNP	chromosome	position	MAF	p-value[¥]	a	var
LS markers						
MARC0043480	10	63867699	0.146	Not significant	0.0220	NA[*]
MARC0006510	11	74240078	0.173	Not significant	-0.0154	NA[*]
E.coli marker						
ALGA0106330	13	145009805	0.388	1e-09	0.5928	2.8%

[¥] *the LS marker were evaluated with the trait number of piglets born alive (NBA) and the E.coli marker with the proportion of animal born that are weaned (PS)*

[*] *The variance explained by the LS markers was not calculated as they did not show any significance regarding the trait NBA*

Table A6: Logarithmic likelihoods of the single-trait GBLUP depending on the statistical model and regarding the traits litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO).

	NBA	UWP	PS	IWO
Model[1]				
GBLUP	-770.50	-3001.80	-3222.61	-3193.02
QTL	-769.13	-3001.80	-3222.06	-3192.02
rQTL	-766.86	-3001.64	-3222.61	-3191.75
LS	-722.33	-3389.96	-3222.89	-3198.17
E.coli	-772.92	-3003.87	-3220.87	-3195.09

[1] *the models are the default GBLUP, the GBLUP with additional QTL-matrix (QTL), the GBLUP with additional rQTL-matrix (rQTL), the GBLUP with additional litter size associated haplotypes (LS) and the model with additional fixed effect of the E.coli marker (E.coli)*

Table A7: Frequency (freq) and effects with standard errors (SE) of the E.coli alleles and of the LS haplotypes regarding the traits litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO).

Allele / haplotype	freq	NBA		UWP		PS		IWO	
		effect	(SE)	effect	(SE)	effect	(SE)	effect	(SE)
E.coli marker									
B	0.39	0	(0.000)	0	(0.000)	0	(0.000)	0	(0.000)
A	0.61	0.054	(0.049)	-0.04	(0.119)	-0.364	(0.128)	0.028	(0.147)
LS haplotypes									
BB	0.7	0	(0.000)	0	(0.000)	0	(0.000)	0	(0.000)
BA	0.12	-0.002	(0.049)	0.002	(0.107)	-0.065	(0.121)	0.074	(0.119)
AB	0.15	0.007	(0.061)	0.276	(0.106)	0.065	(0.148)	-0.077	(0.143)
AA	0.02	-0.054	(0.096)	0.014	(0.219)	-0.071	(0.239)	0.102	(0.235)

Table A8: The prediction accuracies for all models, such as the default GBLUP (GBLUP), the GBLUP with additional QTL-matrix (QTL), the GBLUP with additional reproduction QTL-matrix (rQTL), the GBLUP with additional litter size associated markers (LS) and the GBLUP with additional fixed effect of the E.coli marker (E.coli) for the 400 masked animals.

	GBLUP	QTL	rQTL	LS	*E.coli*
Trait[1]					
NBA	1.147	1.148	1.149	1.147	1.147
UWP	1.158	1.158	1.159	1.158	1.158
PS	1.100	1.100	1.100	1.099	1.099
IWO	1.110	1.109	1.111	1.109	1.110

[1] *the traits concerned are litter size (NBA), proportion of underweighted piglets (UWP), survival rate (PS) and interval weaning to oestrus (IWO)*

Norges miljø- og biovitenskapelig universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway