Exploiting genomic information on purebred and crossbred pigs

ISSN: 1652-6880
ISBN (print version): 978-91-576-8438-7
ISBN (electronic version): 978-91-576-8439-4
ISBN: 678-94-6257-601-8

Thesis committee

Promotor
Prof. Dr M.A.M. Groenen
Personal chair at the Animal Breeding and Genomics Centre
Wageningen University

Thesis co-promotors
Prof. Dr D.J. de Koning
Professor in Animal Breeding at the Department of Animal Breeding and Genetics
Swedish University of Agricultural Sciences

Dr J.W.M. Bastiaansen
Researcher, Animal Breeding and Genomic Centre
Wageningen University

Other members (assessment committee)
Prof. Dr B. Kemp, Wageningen University
Prof. Dr F.A. van Eeuwijk, Wageningen University
Prof. Dr L. Rydhmer, Swedish University of Agricultural Sciences, Sweden
Prof. Dr P. Uimari, University of Helsinki, Finland

The research presented in this doctoral thesis was conducted under the joint auspices of the Swedish University of Agricultural Sciences and the Graduate School of Wageningen Institute of Animal Sciences of Wageningen University and is part of the Erasmus Mundus Joint Doctorate Program “EGS-ABG”.

Exploiting genomic information on purebred and crossbred pigs

André Marubayashi Hidalgo

ACTA UNIVERSITATIS AGRICULTURAE SUECIAE DOCTORAL THESIS No 2015:120

Thesis

submitted in fulfillment of the requirements for the degree of doctor from
Swedish University of Agricultural Sciences
by the authority of the Board of the Faculty of Veterinary Medicine and Animal Science and from
Wageningen University
by the authority of the Rector Magnificus, Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board of Wageningen University and the Board of the Faculty of Veterinary Medicine and Animal Science at
the Swedish University of Agricultural Sciences
to be defended in public
on Wednesday December 9, 2015
at 4 p.m. in the Aula of Wageningen University

Hidalgo, A.M.
Exploiting genomic information on purebred and crossbred pigs

Joint PhD thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden and
Wageningen University, Wageningen, the Netherlands (2015)
With references, with summary in English

Abstract

Hidalgo, A.M. (2015). Exploiting genomic information on purebred and crossbred pigs. Joint PhD thesis, between Swedish University of Agricultural Sciences, Sweden and Wageningen University, the Netherlands

The use of genomic information has become increasingly important in a breeding program. In a pig breeding program, where the final goal is an increased crossbred (CB) performance, the use of genomic information needs to be thoroughly evaluated as it may require a different strategy of what is applied in purebred (PB) breeding programs. In this thesis, I explore the use of genomic information for the genetic improvement of PB and CB pigs. I first focus on the identification of genomic regions affecting traits that are important to breeders. I identified two quantitative trait loci (QTL) regions for gestation length, one for Dutch Landrace on *Sus scrofa* chromosome (SSC) 2 and the other one for Large White on SSC5. I also fine-mapped and narrowed down the region of a previously detected QTL for androstenone level SSC6 from 3.75 Mbp to 1.94 Mbp. A tag-SNP of this fine-mapped region was further investigated and no unfavorable pleiotropic effects were found; indicating that using the studied marker for selection would not unfavorably affect the other studied traits. After that, the focus was changed to the application of genomic selection in pigs. Within-population predictions showed high accuracies, whereas across-population prediction had accuracies close to zero. Using combinations among Dutch Landrace and Large White populations plus their cross showed that multi-population prediction was not better than within-population. The exception was when the CB pigs were predicted with records from both parental populations added to the CB training data. When using PB pigs to train CB ones, the predictive ability found indicates that selection in the PB pigs results in response in the CB ones. When assessing the source of information used to estimate the breeding values used as response variable, I showed that a more accurate prediction of CB genetic merit was found when training on PB data with breeding values estimated using CB performance than training on PB data with breeding values estimated using PB performance. I also studied the accuracy of using CB pigs in the training population to select PB for CB performance. Predictive ability when using CB phenotypes for training was observed, however, the accuracy was lower than using PB phenotypes in the training population. Lastly, I evaluate the inclusion of dominance in the model when using a CB training population. Results showed that accounting for dominance effects can be slightly beneficial for genomic prediction compared with a model that accounts only for additive effects.

List of Publications

This thesis is based on the work contained in the following papers:

I – Hidalgo, A.M., Lopes, M.S., Harlizius, B. and Bastiaansen, J.W.M. Genome-wide association study reveals regions associated with gestation length in two pig populations. *Animal Genetics (accepted).*

II – Hidalgo, A.M., Bastiaansen, J.W.M., Harlizius, B., Megens, H.J., Madsen, O., Crooijmans, R.P.M.A. and Groenen, M.A.M. (2014). On the relationship between an Asian haplotype on chromosome 6 that reduces androstenone levels in boars and the differential expression of *SULT2A1* in the testis. *BMC Genetics* 15:4.

III – Hidalgo, A.M., Bastiaansen, J.W.M., Harlizius, B., Knol, E.F., Lopes, M.S., De Koning, D.J. and Groenen, M.A.M. (2014). Asian low-androstenone haplotype on pig chromosome 6 does not unfavorably affect production and reproduction traits. *Animal Genetics* 45(6), pp. 874-877.

IV – Hidalgo, A.M., Bastiaansen, J.W.M., Lopes, M.S., Harlizius, B., Groenen, M.A.M. and De Koning, D.J. (2015). Accuracy of predicted genomic breeding values in purebred and crossbred pigs. *G3: Genes | Genomes | Genetics* 5, 1575-1583.

V – Hidalgo, A.M., Bastiaansen, J.W.M., Lopes, M.S., Veroneze, R., Groenen, M.A.M. and De Koning, D.J. (2015). Accuracy of genomic prediction using deregressed breeding values estimated from purebred and crossbred offspring phenotypes in pigs. *Journal of Animal Science* 93(7), pp. 3313-3321.

VI – Hidalgo, A.M., Bastiaansen, J.W.M., Lopes, M.S., Calus, M.P.L. and De Koning, D.J. Accuracy of genomic prediction of purebreds for crossbred performance in pigs (manuscript).

VII – Hidalgo, A.M., Zeng, J., Fernando, R.L., Lopes, M.S. and Dekkers, J.C.M. Evaluation of genomic prediction of purebreds for crossbred performance in pigs accounting for dominance effects (manuscript).

1

General introduction

1.1 Introduction

Animal breeding aims to select the best animals to be the parents of the next generation. A large variety of techniques, strategies and methods have been developed to achieve this goal. In recent years, genotyping technology has improved considerably and high-throughput genomic information became available. Efficient use of this information, hence, is crucial for the competitiveness of a breeding company. In this work, therefore, I will explore the use of genomic information for the genetic improvement of purebred and crossbred pigs. In this general introduction, I will first concentrate on the identification of genomic regions that affect traits that are important to breeders. After that, I will focus on the application of genomic selection, and later on crossbreeding with emphasis on heterosis and dominance. These topics are relevant in the application of genomic information in the present breeding situation.

1.2 QTL mapping

Most traits of economic importance in livestock production are quantitative, i.e., are affected by many loci to various degrees. The genes affecting a quantitative trait, so-called "quantitative trait loci" (QTL), are difficult to identify, yet they are relevant for breeding purposes. Currently, 13,030 QTL for 663 traits have been described for pig (Animal QTLdb, http://www.animalgenome.org/QTLdb).

Genetic markers can be divided in three groups: 1) direct markers: loci that code for the causative mutation, 2) LD markers: loci are in population-wide linkage disequilibrium with the causative mutation, 3) LE markers: loci are in population-wide linkage equilibrium with the causative mutation in outbred populations (Dekkers 2004). Direct markers are the most difficult to detect because proving causality is extremely hard. The LD markers can be detected using candidate genes (Rothschild and Soller 1997), fine-mapping (Andersson 2001; Georges 2007) or genome-wide association studies (GWAS); LD markers are located close to the causative mutation so that linkage disequilibrium between marker and QTL exists. The LE markers within linkage distance of a QTL can be identified by using breed crosses or analysis of large half-sib families within the breed.

The first study that detected a QTL in pigs, identified a region affecting fat deposition on chromosome 4 (Andersson *et al.* 1994). This study, along with other contemporaneous studies, performed linkage mapping in an F2 design using microsatellite markers spread across the genome. The F2 were, in general, obtained from crosses between a European-descent commercial breed and either a European Wild Boar or Asian breed, such as Meishan (e.g. Knott *et al.*, 1998; De Koning *et al.*, 1999). Many QTL regions were detected using this methodology (reviewed by Rothschild *et al.* (2007)), however the confidence interval of these QTL were usually very large which hampered the use of this information in a breeding program. On top of the large confidence intervals, most of these QTL were detected in experimental populations using crosses, therefore the identified QTL could hardly be used directly for selection within breeds as they differ in frequency across breeds (Dekkers 2004). In practice, QTL analysis in crossed populations has been superseded by GWAS analyses within purebred populations, which will be described later.

The fine-mapping approach aims to find the causative mutation or at least refine the mapping resolution of a previously detected QTL region, which should lead to narrowing down this QTL region. The major factors affecting the mapping resolution are: 1) marker density, 2) crossover density, 3) accuracy of inferring the QTL genotype, and 4) molecular architecture of the QTL (Georges 2007). Provided that there are enough markers, then to increase the mapping resolution, there is the need to increase the number of recombinations. This increase can be achieved by breeding additional generations or increasing the population size (Darvasi and Soller 1995). The fine-mapping approach has been successful in detecting the causal mutation only for a small number of QTL, for example *FAT1* (Berg *et al.* 2006) and the insulin-like growth factor 2 gene (*IGF2*) (Van Laere *et al.* 2003).

Besides the linkage approach used for QTL mapping, other approaches were developed and applied in pig breeding, such as the candidate gene approach. The candidate gene approach involves 1) selecting the candidate gene based on its known biological function, 2) amplifying the gene, 3) finding polymorphic regions, 4) large scale genotyping of the polymorphic region, 5) phenotyping and genotyping a target population, 6) performing an association between phenotype and genotype, and finally 7) assessing the detected associations (Rothschild and Soller 1997). The candidate gene approach was successful in detecting few QTL, for example the porcine melanocortin-4 receptor (*MC4R*) gene (Kim *et al.* 2000). This approach discovered LD markers, which allows selection across animals of the same population, therefore is relevant for breeding (Dekkers 2004).

The pig genome sequence was published in 2012 by the Swine Genome Sequencing Consortium (Groenen *et al.* 2012). In the meantime, the identification of high numbers of single nucleotide polymorphisms (SNP) and the development of methodologies to simultaneously genotype large numbers of SNP, enabled the design of a SNP chip for pigs with approximately 60,000 markers (Ramos *et al.* 2009). The higher marker density across the genome allowed performance of genome-wide association mapping, for the identification of QTL. GWAS evaluates whether variations in the genome (e.g. SNP) are associated with variation in a given trait. The assumption underlying a GWAS is that significant associations occur because the SNP is in linkage disequilibrium (LD) with a causative mutation affecting the trait. The first study performing a GWAS in pigs identified a cluster of markers associated with androstenone level on chromosome 6 (Duijvesteijn *et al.* 2010).

To make use of markers linked to QTL in breeding, Fernando and Grossman (1989) developed a methodology that incorporated markers associated with quantitative traits into the conventional mixed models genetic evaluation. This method was applied by breeding companies as a complementary tool to the pedigree-based genetic evaluation (Ibáñez-Escriche *et al.* 2014). Before incorporating new markers in the genetic evaluation, it is recommended to assess the pleiotropic effects of that marker on other production and reproduction traits. This check is important to avoid unfavorable effects due to pleiotropy and/or due to genetic hitchhiking. Such unfavorable effects are examined by testing the association between the marker and the other traits.

So far, only a handful of causative mutations has been discovered and for the majority of QTL regions the causal variation has not been identified. The general finding from GWAS for quantitative traits, in livestock species, is that the majority of the economically important traits are controlled by many genes with small effects. Therefore, given the polygenic nature of most traits in livestock and the availability of a large number of genetic markers across the genome, genomic selection became the method of choice for application in animal breeding.

1.3 Genomic selection

Genomic selection (GS) entails using markers across the genome to estimate breeding values (Meuwissen *et al.* 2001). The assumption underlying genomic selection is that the effects of QTL will be captured by markers due to LD. In GS, individuals with both phenotypes and genotypes compose the so-called training population. Information on the training population is used to estimate direct genomic values (DGV) of selection candidates that are genotyped but do not have phenotypes. Selection based on DGV can be performed in these selection candidates. The DGV is an estimate, based on the animal's genomic information, of the value that an animal transfers to its progeny. To calculate the DGV, marker effects can be estimated by regressing the phenotypes on the marker genotypes in the training population. Afterwards, the genotypes of each selection candidate are multiplied by the marker effect and summed, resulting in the DGV. Various methods have been developed for the application of GS. These methods are generally based on mixed models, simple linear regression or shrinkage-based approaches. A detailed overview and evolution of these methods is described by Garrick *et al.* (2014).

In animal breeding, the selection of the best animals to be the parents of the next generation is performed typically to achieve a response to selection. The response to selection (*R*) is determined by the intensity of selection (*i*), the accuracy of prediction (*r*), the genetic standard deviation (σ_a) and the generation interval (*L*):

$$R = \frac{i * r * \sigma_a}{L}$$

Studies on genomic predictions have shown a solid increase in accuracy over pedigree-based predictions (BLUP). The degree of increase varies across traits, lines and species (e.g. Hayes *et al.*, 2009; Tussel *et al.*, 2013). In addition to the increase in accuracy, GS allows selection at a younger age of the selection candidates because the genotype that will be used for prediction can be obtained right after birth. Therefore, there is no need to spend a long time waiting for the expression and recording of the animals own phenotype, e.g. daily gain, or the phenotype of their offspring, e.g. milk production. This leads to a reduction in the generation interval, which is a larger benefit in some species (e.g. cattle) than in others (e.g. broilers). The potential for changing the intensity of selection with GS exists but it depends on the number of genotyped individuals; the more genotyped animals the higher the intensity and therefore a greater expected response to selection. Genomic selection, therefore, can affect response to selection through these three factors, *i*, *r* and *L*.

Genomic selection was first applied in dairy cattle (VanRaden *et al.* 2009), where the aim is to improve the performance of purebred animals. In pigs, two major pig breeding companies (PIC, Topigs Norsvin) began GS implementation in purebred lines in 2012-13. The delay in implementing GS in pigs, compared to cattle, can be attributed to: 1) the later release of the commercial SNP chip (Jan. 2008 for cattle vs Aug. 2009 for pigs), 2) the high genotyping cost compared to the value of an animal,

3) the different structure of the business (open nucleus vs. closed nucleus), 4) the need to distinguish from competitors in the market, 5) the uncertainty whether GS of purebreds results in gains in the crossbreds. The latter (crossbred production) plays an important role in pig production, and the crossbred breeding goals in pigs is probably a main difference between dairy cattle breeding and pig breeding. Implementation of GS in pigs for the crossbred breeding goals, hence, may require different strategies which are not yet fully developed. Besides the different strategies that need to be assessed, the accuracy of methods that are currently implemented for cattle may be reduced when the aim is to improve crossbred performance. Many factors affect this lower accuracy, such as the low number of genotyped crossbred individuals, genetic correlation between purebred and crossbred performance being different from 1, and the lower relationship between the purebred and crossbred individuals. Assessing accuracy of genomic prediction for the performance of purebred and crossbred animals, therefore, is a research field in development and of great interest for pig and poultry breeding companies.

1.4 Crossbreeding

Crossbreeding is the process of mating individuals from different breeds or lines to produce a crossbred offspring. It is standard practice in the modern pig production set-up, and as indicated in the preceding section, is a relevant difference compared to, for instance, dairy cattle breeding. Crossbreeding is applied to capitalize on breed complementarity and heterosis, and to protect the genetic progress in the pure lines.

Focusing on the importance of heterosis for crossbreeding, three types can be distinguished: individual, maternal and paternal (Clutter 2010). It is the individual heterosis that benefits the crossbred progeny and is a result of its own hybrid state and the primary aim for improving production traits. The maternal heterosis benefits the crossbred progeny and is a result of the hybrid state of its dam. Maternal heterosis is highly relevant for reproduction traits, e.g. mothering ability, because it benefits the offspring especially in the period that the offspring is dependent on its dam. Maternal heterosis is therefore a major reason for the extensive use of two-generation crossbreeding schemes in pig production (Bidanel 2010). The paternal heterosis benefits the crossbred progeny and is a result of the hybrid state of its sire. The benefit of paternal heterosis is limited, not having the same relevance as the maternal heterosis in crossbreeding. In general, heterosis is found across traits and species and varies roughly from 0% to 30%, including negative values as well (Bondoc *et al.* 2001; Bidanel 2010).

Dominance is labelled to be one of the main causes of heterosis (Falconer and Mackay 1996; Charlesworth and Willis 2009). This is because the hybrid superiority is attributed to the advantage of the heterozygotes over the mean of the two homozygotes. Studies in pigs and cattle have found that there is dominance variance for different traits in purebred populations (Su *et al.* 2012; Nishio and Satoh 2014; Sun *et al.* 2014). In addition, these studies have also reported that using a model that accounts for dominance resulted in either higher or similar accuracy for prediction of breeding values than using a model that only fits additive effects. Prediction of crossbred performance, accounting for dominance, has not been reported. Accounting for dominance in prediction of crossbreds is expected to result in a considerable increase of accuracy compared to purebred results because more dominance is envisaged in crossbred than purebred populations (Nishio and Satoh 2014). Therefore, using a model that accounts for dominance when crossbred individuals are used in the prediction might be important.

1.5 This thesis

The objective of my research is to exploit genomic information in purebred and crossbred pigs to generate knowledge and results that could be used to improve genetic progress. The thesis can be divided in two parts: 1) in this part the aim is to discover and investigate genomic regions that affect gestation length and boar taint, including an assessment of pleiotropic effects of the identified marker; 2) in this part the potential of genomic selection in pig breeding is investigated by determining the accuracy of genomic prediction using different training and validation populations, selected from multiple purebred lines and their crossbred offspring, and different models.

The first part of this thesis comprises Chapters 2-4 and concentrates on finding important genomic regions and test for possible application of these results in pig breeding. In **Chapter 2**, a GWAS is described with the aim to detect SNP and also to identify candidate genes that are associated with gestation length. Gestation length is an important trait in pig breeding due to its relation with maturity of the piglet at birth. Detecting significant SNP with effects on gestation length is therefore desired. In **Chapter 3**, the region of a previously detected QTL is fine-mapped, aiming at the identification of SNP that affect androstenone levels. This fine-mapped region is evaluated in **Chapter 4** for possible pleiotropic effects on production and reproduction traits in pigs. The combined results of Chapters 3 and 4 allow an informed decision on the usage of these markers in a breeding program.

The second part comprises Chapters 5-8 and focuses on strategies to implement GS in pig breeding when crossbreeding schemes are accounted for. In **Chapter 5**, the accuracy of genomic breeding values from within-, multi- and across-population predictions in pigs is evaluated, including the accuracy of using purebred training data to predict performance of crossbred pigs. This last analysis will indicate how well crossbred performance will respond to the current practice of selecting within purebred populations. For this chapter, the response variable used for training was the deregressed breeding value (DEBV) from a routine genetic evaluation, which contains a mix of purebred and crossbred animals. To separately assess the value of phenotypic information from purebred and crossbred pigs I investigated the source of information used to estimate the DEBV: should it be based on purebred or crossbred performance? Therefore, in **Chapter 6**, while the training and validation populations were the same as in Chapter 5, the training was performed twice with different phenotypes as input: first using DEBV based on purebred offspring, and second using DEBV based on crossbred offspring. The DEBV from crossbred offspring is expected to lead to better predictions of purebred animals for crossbred offspring performance. Later, more genotyped crossbred animals became available and a training population could be constructed that consisted of genotyped crossbred animals. Hence, in **Chapter 7** we compare the accuracy of prediction from using either only crossbred or only purebred animals as training population when predicting purebred animals for crossbred performance. Finally, as indicated above, the performance of crossbreds typically shows heterosis, and dominance is expected to strongly contribute to this heterosis. Therefore in **Chapter 8**, the performance of the dominance model is empirically compared to the additive model for prediction of purebreds for crossbred performance based on a training with data from crossbred pigs.

Lastly, **Chapter 9** is where the two parts, mapping and prediction, come together. I discuss the relevance of my findings, how breeders can benefit from the combination of genomic selection with the information of important genomic regions identified in GWAS. Also, I discuss the impact that high-density SNP chips and sequence data can have in GWAS studies. In addition, I expatiate on strategies for applying genomic selection, especially when crossbreeding information is used. To finalize, I give concluding remarks by summarizing the new insights from this thesis.

References

Andersson L (2001) Genetic dissection of phenotypic diversity in farm animals. Nat Rev Genet 2:130–138.

Andersson L, Haley CS, Ellegren H, et al (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science 263:1771–1774.

Animal QTLdb. http://www.animalgenome.org/QTLdb. Accessed 21 Aug 2015

Berg F, Stern S, Andersson K, et al (2006) Refined localization of the FAT1 quantitative trait locus on pig chromosome 4 by marker-assisted backcrossing. BMC Genet 7:17.

Bidanel J (2010) Biology and genetics of reproduction. In: Rothschild MF (ed) The genetics of the pig. CABI, pp 218–233

Bondoc OL, Santiago CAT, Tec JDP (2001) Least-square analysis of published heterosis estimates in farm animals. Philipp J Vet Anim Sci 27:12–26.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nat Rev Genet 10:783–96.

Clutter AC (2010) Genetics of Performance Traits. In: Rothschild MF (ed) The genetics of the pig. CABI, pp 325–348

Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199–1207.

De Koning DJ, Janss LL, Rattink AP, et al (1999) Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*). Genetics 152:1679–1690.

Dekkers JC (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci 82:E313–E328.

Duijvesteijn N, Knol EF, Merks JWM, et al (2010) A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genet 11:42.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, 4th edn. Longman, Harlow, England

Fernando RL, Grossman M (1989) Marker assisted selection unbiased prediction using best linear. Genet Sel Evol 21:467–477.

Garrick D, Dekkers J, Fernando R (2014) The evolution of methodologies for genomic prediction. Livest Sci 166:10–18.

Georges M (2007) Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. Annu Rev Genomics Hum Genet 8:131–62.

Groenen MAM, Archibald AL, Uenishi H, et al (2012) Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491:393–8.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–43.

Ibáñez-Escriche N, Forni S, Noguera JL, Varona L (2014) Genomic information in pig breeding: Science meets industry needs. Livest Sci 166:94–100.

Kim KS, Larsen N, Short T, et al (2000) A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. Mamm Genome 11:131–135.

Knott SA, Marklund L, Haley CS, et al (1998) Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. Genetics 149:1069–80.

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–29.

Nishio M, Satoh M (2014) Including dominance effects in the genomic BLUP method for genomic evaluation. PLoS One 9:e85792.

Ramos AM, Crooijmans RPMA, Affara NA, et al (2009) Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One 4:e6524.

Rothschild MF, Hu Z, Jiang Z (2007) Advances in QTL mapping in pigs. Int J Biol Sci 3:192–7.

Rothschild MF, Soller M (1997) Candidate gene analysis to detect genes controlling traits of economic importance in domestic livestock. Probe (Lond) 8:13–20.

Su G, Christensen OF, Ostersen T, et al (2012) Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One 7:e45293.

Sun C, VanRaden PM, Cole JB, O'Connell JR (2014) Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. PLoS One 9:e103934.

Tusell L, Pérez-Rodriguez P, Forni S, et al (2013) Genome-enabled methods for predicting litter size in pigs : a comparison. Animal 7:1739–1749.

Van Laere A-S, Nguyen M, Braunschweig M, et al (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425:832–836.

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24.

2

General discussion

2.1 Introduction

With the development of high-throughput and cost-effective genotyping methods, exploiting genomic information became an indispensable approach for major breeding companies. Pig production relies on crossbreeding, hence, the use of genomic data for selection for crossbred performance needs to be carefully assessed. Implementation of genomic selection in crossbreeding schemes cannot be a simple copy of what is applied in breeding programs for purebred performance.

For the research presented in this thesis, I used genomic information from purebred and crossbred pigs. I have detected genomic regions associated with gestation length and with androstenone level by genome-wide association and fine-mapping analyses. Further, I studied potential pleiotropic effects of the androstenone level QTL on chromosome 6 on production and reproduction traits. To investigate the potential and peculiarities of applying genomic selection in a crossbreeding setting, I evaluated and showed that there is predictive ability between purebred and crossbred pigs. Consequently, genomic selection in purebred pigs will result in gains in the performance of crossbreds. In this Chapter, I discuss the relevance of my findings in a broader context. I will discuss how to integrate individual genetic markers with genomic selection, as well as different strategies for applying genomic selection in pig breeding using genotypes and phenotypes of purebred and crossbred animals.

2.2 Integrating individual genetic markers with genomic selection

For qualitative traits, DNA tests were developed, starting some 25 years ago, which allowed selection against an undesired condition or phenotype. For example, a recessive allele (HAL 1843$^{\text{TM}}$) in the porcine ryanodine receptor (*RYR1*) gene that causes malignant hyperthermia in stressful conditions (Fujii *et al.* 1991). When a single locus is controlling the trait, a DNA test is an effective tool for selection. The majority of the production traits in livestock, however, are continuously distributed (quantitative) because many quantitative trait loci (QTL) are controlling the trait. Due to the high number of loci affecting the trait, individual QTL only explain a proportion of the total genetic variance.

Because of the typically small effects, selection based only on individual markers was not applied in pig breeding companies. This was in contrast with the expectations that were set after the initial boom of genetic markers (Ibáñez-Escriche *et al.* 2014).

Genetic markers that explain part of the variance and are in linkage disequilibrium with a QTL, were incorporated into the genetic evaluation using customized SNP panels (Van Eenennaam *et al.* 2014). Such markers were used as complementary tool (Ibáñez-Escriche *et al.* 2014) resulting in marker-assisted BLUP (MA-BLUP) being applied by pig breeding companies. Like most QTL, the QTL regions for gestation length identified in Chapter 2 also explained a relatively small proportion of the genetic variance, 1.12% for the Dutch Landrace and 0.77% for the Large White pigs. Further, in Chapter 3, I fine-mapped a previously identified QTL region for androstenone level that also explained a small proportion of phenotypic variance, 6% in the Duroc population (Duijvesteijn *et al.* 2010). These results are concordant with the vast literature that reported 13,030 QTL for 663 traits usually with small effects (Animal QTLdb, http://www.animalgenome.org/QTLdb).

With the development of methods that allow to perform genomic prediction based on a large number of genetic markers (Meuwissen *et al.* 2001), and after the availability of commercial SNP chips, genomic selection (GS) became the center of attention for animal and plant breeders. Since then, GS has been implemented in dairy cattle (VanRaden *et al.* 2009) and it was shown to result in higher accuracies than traditional genetic evaluations (BLUP) (Hayes *et al.* 2009b). The main positive point of GS lies in its ability to capture the infinitesimal nature of the majority of economically important traits, which was exactly the main cause for the limited success of marker-assisted selection. In GS, all markers have their effects estimated without the need to know the biological meaning. All that is needed is a training population and sufficient computational power to run the genomic evaluation. The training population, which is phenotyped and genotyped, has to have sufficient size (Misztal 2011) and preferably be related to the selection candidates.

Even though only few causative mutations have been identified so far, such significant markers will continue to be identified. Further developments in genotyping technology resulted in a reduction of costs, enabling the production of commercial high-density (HD) SNP chips (e.g. Illumina Bovine HD 770k SNP chip). Therefore, with more animals genotyped, which increases the sample size, and with the genome more densely covered with markers, which leads to a smaller distance between the SNP and the causative mutation, a more precise detection of QTL is expected. Genome-wide association studies (GWAS) using HD SNP panels have been performed in cattle (e.g. Purfield *et al.* (2015)). In pigs, a HD SNP chip has been recently developed with approximately 660,000 SNP, however, GWAS with this HD SNP chip are still lacking. The ultimate level of genotypic information is the sequence data. Sequencing determines the order of all nucleotides of the DNA of a given organism. Therefore, sequence data contain the causative mutations of the trait. A GWAS using sequence data, hence, is expected to find the causative mutation (Meuwissen and Goddard 2010). There have been efforts to increase the numbers of sequenced animals (e.g. Daetwyler *et al.* (2014)), to enable GWAS with sequenced individuals. The approach that has been taken is to perform a GWAS using HD SNP chip genotype data and then focus on the identified peaks, performing a region-wise association study (RWAS) using imputed sequence data (Sahana *et al.* 2014; Wu *et al.* 2015). This method was able to refine previously detected QTL regions, however, it was not able to identify the causative mutation, mainly because of strong blocks of linkage disequilibrium. Another factor that might be hampering the identification of the causative mutation is that imputation is not 100% accurate, especially for rare variants and small reference panels.

As these significant regions on the genome are still being found and described, it is of interest to integrate the significant markers in the genomic evaluation. This integration is relevant because, while the causative mutations are not detected, these significant markers provide knowledge regarding the genetic architecture of the trait. Although the effects found are not large, they might add to the prediction accuracy and thus should be explored. Integrating these markers into the genomic evaluation would be a form of marker-assisted genomic prediction. Here, the marker genotype (0, 1 or 2) is fitted as a fixed effect in the genomic prediction model (MA-GBLUP). The outcome of this analysis is an estimate of estimated breeding value (EBV) of the animal and an estimate of the marker's allele substitution effect. After that, multiplying the estimate of the marker effect by the animal’s genotype (0, 1 or 2) and adding this value to the EBV results in the animal’s EBV from MA-GBLUP. MA-GBLUP offers the possibility to apply the results described in Chapters 2 and 3 to within-population genomic predictions as described in Chapters 5-7.

Before implementing MA-(G)BLUP it is important to know the effect of the QTL on all traits in the breeding goal. Hence, assessing pleiotropic effects of that marker on other traits is recommended to avoid unfavorable effects due to pleiotropy and/or due to genetic hitchhiking. Grindflek *et al.* (2011) found markers on the pig genome affecting simultaneously the levels of boar taint compounds (e.g. androstenone) and of sex hormones. Given that the androstenone markers have an unfavorable impact on sex hormones, the use of such markers for selection would be challenging. I showed in Chapter 4, however, that selection for the marker on chromosome 6 that reduces androstenone level will have no unfavorable effect on production and reproduction traits studied. Therefore, the use of that marker to reduce androstenone level in a breeding program becomes of interest.

To show whether integrating significant markers with genomic prediction is relevant, I performed a MA-GBLUP analysis using the most significant marker of each population described in Chapter 2 and the marker studied in Chapter 4. Markers were: rs81308021 for androstenone level in the Duroc, rs81366467 for gestation length in the Dutch Landrace and rs344547786 for gestation length in the Large White. Individuals from three pig populations were used: 833 Duroc, 1,615 Dutch Landrace and 1,904 Large White animals. These animals were genotyped using the Illumina PorcineSNP60 BeadChip (Ramos *et al.* 2009) and quality control was performed on the genotypes according to the methods described in Chapter 5. After quality control, 41,289 SNP for the Duroc, 42,360 SNP for the Dutch Landrace and 41,005 SNP for the Large White remained out of the initial 64,232 SNP. We analysed the data using ASReml 3.0 (Gilmour *et al.* 2009) with the model:

$$\mathbf{y} = \mu + \mathbf{b}_1\mathbf{SNP} + \mathbf{Zu} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of pre-corrected phenotypes, μ is the overall mean, $\mathbf{b}_1$ is the vector of regression coefficients of each SNP, $\mathbf{SNP}$ is the incidence vector for $\mathbf{b}_1$ with genotypic information (0, 1 and 2), $\mathbf{Z}$ is the incidence matrix for $\mathbf{u}$, $\mathbf{u}$ is the vector of random additive genetic effects, assumed to be $\sim N(\mathbf{0}, \mathbf{G}\sigma_u^2)$, where $\mathbf{G}$ is the genomic relationship matrix, and e is the residual error, assumed to be $\sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is an identity matrix. The accuracy of prediction was estimated as the correlation between the EBV and the corrected phenotype in a set of validation animals. The validation population consisted of the 20% youngest genotyped animals of a given population. Phenotypes were corrected for fixed effects as described in Chapter 5. Prediction results of MA-GBLUP were compared to the results obtained from using the traditional genetic evaluation (BLUP), marker-assisted BLUP (MA-BLUP) and genomic evaluation (GBLUP) (Table 2.1).

Table 2.1 Accuracies of prediction for androstenone level (AND) and gestation length (GLE) using different methods.

Trait	Breed	$N_{training}$	$N_{validation}$	Accuracy† (Bias*)			
				BLUP	MA-BLUP	GBLUP	MA-GBLUP
AND	DU	666	167	0.39 (1.43)	0.42 (1.29)	0.43 (1.01)	0.45 (1.07)
GLE	DL	1,292	323	0.29 (0.73)	0.31 (0.79)	0.41 (0.81)	0.42 (0.81)
GLE	LW	1,523	381	0.41 (1.11)	0.41 (1.11)	0.46 (0.90)	0.46 (0.90)

DU - Duroc, DL - Dutch Landrace, LW - Large White, N - number of animals

† - Correlation between EBV and pre-corrected phenotype

* - Regression coefficient of the phenotype on the EBV

MA-GBLUP resulted in the highest accuracy for all three analyses (Table 2.1). In the Large White population, no difference was observed from either including or excluding the marker as fixed effect for gestation length when comparing BLUP with MA-BLUP, nor when comparing GBLUP with MA-GBLUP. This result in the Large White population is probably due to the minor allele frequency (MAF) of the most significant marker being very low (0.01) (Chapter 2), which means that the majority of the animals had the same genotype. Therefore adding the same marker effect to the EBV of the vast majority of the animals would not affect the accuracy. For androstenone level in the Duroc, and for gestation length in the Dutch Landrace, there was an increase in accuracy when the significant marker information was used. The increase in accuracy for MA-BLUP over BLUP was greater than for MA-GBLUP over GBLUP. As BLUP uses only pedigree information, fitting the most significant marker as fixed effect can differentiate animals with regard to the QTL, leading to a possible increase in accuracy. The increase in accuracy of MA-GBLUP over GBLUP was not as great because GBLUP already accounts for the significant marker in the **G** matrix. However, even when the same genotypic information is present in the **G** matrix, fitting the significant marker separately as a fixed effect still resulted in higher accuracy of prediction because the marker effect is better captured by the model. Fitting the marker as a separate fixed effect is not expected to lead to lower accuracies, even if the marker is a false-positive. In such a case, the effect estimated would be zero, accuracy would remain the same, and thus no harm would be done to the prediction. An issue will occur when trying to fit more markers as fixed effects than the number of animals. In this case, estimation problems occur because of a lack of degrees of freedom to fit all effects simultaneously by least squares (Lande and Thompson 1990). However, markers with large effects are not so common, therefore this issue is not likely to become a problem for the MA-GBLUP model. The regression coefficients of the phenotype on the EBV were in general close to 1 in all analyses included in Table 2.1, which indicates unbiased predictions. Less bias was observed for MA-BLUP than for BLUP, and for the genomic models GBLUP and MA-GBLUP compared with MA-BLUP and BLUP. These analyses were performed in purebred animals, therefore I can predict that MA-GBLUP would result in greater response to selection in the pure lines over GBLUP. In a breeding program where the goal is to select purebred animals for purebred performance, MA-GBLUP is therefore recommended for traits with known significant marker(s). To extrapolate to prediction of crossbred performance, MA-GBLUP would be beneficial for both purebred and crossbred performance when the QTL is the same for purebred and crossbred performance. If the interest is to select purebred animals for crossbred performance, as is the case in pig breeding, I would expect that using MA-GBLUP could improve accuracy of prediction as long as the marker is affecting the crossbred population.

2.3 Genomic selection in pigs

Genomic selection has been introduced in dairy cattle breeding aiming to improve performance of purebred animals (VanRaden *et al.* 2009). In pigs, however, the end product is a crossbred animal which may require different strategies for the implementation of GS from what is currently applied in dairy cattle. In pig breeding, specialized sire and dam lines are kept in the breeding stock and crossed to produce a three-way or four-way cross finisher pig (Merks and De Vries 2002).

In this thesis, I have analyzed androstenone level and reproduction traits. Reproduction traits generally have low heritability, but gestation length has moderate heritability. Genomic selection has a large added value for low-heritability traits (Muir 2007; Calus *et al.* 2008) because the accuracies of these traits are usually low as they depend on the heritability of the trait (Falconer and Mackay 1996; Muir 2007; Visscher *et al.* 2008). For production traits, which generally have higher heritabilities, traditional genetic evaluation already provides EBV with high accuracy, therefore the added value of GS is less. In addition to heritability, other factors affect the value of GS, e.g. the time at which traits are measured. GS can have a great positive impact on the accuracy of EBV for meat-quality traits, which are measured after slaughter therefore usually measured on relatives of selection candidates. Also, GS is expected to have a larger impact on sex-limited traits, traits that are difficult (expensive) to record, and on traits that are recorded late in life (Muir 2007). This positive impact occurs because the accuracy of traditional genetic evaluation is limited for these traits.

In this section, I will discuss different strategies of genomic selection in pigs and their perspectives. The use of within-, across- and multi-population predictions will be discussed, along with the use of crossbred information for genomic prediction.

2.3.1 Within-population prediction

Pig breeders have focused on the estimation of breeding values of purebred animals using data obtained also from purebred animals which are kept in nucleus farms. In other words, the selection is applied to improve purebred genetic merit with an expectation for a response in crossbreds. In Chapters 5 and 6, results of within-population genomic predictions are presented which showed considerably high accuracy of prediction. Within population, genomic prediction generally performed better than traditional genetic evaluation based on pedigree, which is also observed in other studies in pigs (e.g. Tusell *et al.* (2013)). Therefore genomic prediction, within-population, is recommended when the aim is to increase purebred performance. In practice, breeding companies currently perform within-population genomic prediction by applying the single-step approach (Misztal *et al.* 2009). This approach is preferred by breeding companies because current data sets still contain a large amount of data on phenotyped animals that are not genotyped. With the single-step approach, these records can still be used together with phenotyped and genotyped individuals to estimate the breeding values. Additionally, the pipeline for implementing the single-step approach is similar to the traditional genetic evaluation that was in use previously. The only major change is the replacement of the average numerator relationship matrix (**A** matrix) with an **H** matrix which contains the pedigree-genomic relationships (Legarra *et al.* 2009).

Once within-population genomic prediction is implemented, accounting for the genetic architecture of the trait might be relevant. Weighting the **G** matrix increases the accuracy of prediction (Zhang *et al.* 2010; Tiezzi and Maltecca 2015; Veroneze 2015). A practical problem is accounting for the genetic architecture in genomic evaluations would require a separate analysis for every single trait because a different **G** matrix would have to be built for each trait. To avoid this problem, using the MA-GBLUP methodology, described above, is a way of accounting for the markers with large effect in a multi-trait genomic evaluation without the need of constructing separate **G** matrices for each trait.

2.3.2 Across-population prediction

In pig breeding, multiple dam and sire lines are kept in the breeding stock. It is possible that a training dataset is not available for a specific line or that a design might be desired in which training data would only be produced in some of the lines. In such cases, performing across-population prediction could be a good strategy (Hayes *et al.*, 2009). Across-population prediction involves using population A as training dataset to predict population B. Studies in cattle have shown that training in one population to predict another results in accuracies close to zero (Harris *et al.*, 2008; Hayes *et al.*, 2009; Chen *et al.*, 2015). This low accuracy has been attributed to the different marker-QTL linkage disequilibrium phase across populations (De Roos *et al.* 2009). In pigs, we have also found accuracies close to zero for across-population predictions (Chapter 5). Therefore, under the current circumstances of a low number of animals, genotyped with around 60,000 SNP, I would not recommend across-population prediction. No matter what the reason for the application of across-population prediction would be, constraints in expenses or genomic breeding program design, the results are not encouraging. Instead, I would perform within-population genomic prediction for the line that has a training population and continue the pedigree-based genetic evaluation for the other line. In the future, when more animals are sequenced and possibly more causative mutations are identified, across-population prediction might yield better accuracies.

2.3.3 Multi-population prediction

An alternative to across-population prediction is to have, in the training set, some animals from the same population that will be predicted, and increase the size of the training set by combining populations A and B. The increase in accuracy from multi-population prediction is highly dependent on the relationship between the combined populations (De Roos *et al.* 2009). Many studies on multi-population prediction were performed in dairy cattle and have been reviewed by Lund *et al.* (2014). Generally, there is an increase in accuracy when the same breeds from different countries are combined, whereas this increase is minor when the breeds are only distantly related. Multi-population prediction in pigs, using Dutch Landrace and Large White animals plus the cross between these two populations was performed in Chapter 5. Results showed that adding the other population in the training set did not improve the accuracy compared with within-population prediction. The main reason for that was that the Dutch Landrace and Large White populations are only distantly related. Predicting the F1 cross using a multi-population training data set, which contained the F1 cross plus both parental populations, was advantageous over within-population prediction when genetic correlation between purebred and crossbred performance was high (>0.9). The parental populations are closely related to the F1 which appears to have a positive impact on accuracy of multi-population prediction (Chapter 5). Also, having a high genetic correlation between purebred and crossbred performance is relevant in boosting the accuracy of multi-population prediction. Thus, multi-population prediction in pig breeding can be recommended when predicting crossbred animals, given that populations are closely related and/or the genetic correlation between purebred and crossbred performance is 1 or close to unity.

2.3.4 Using crossbred information for genomic prediction

The final goal in pig breeding is to improve performance of the commercial crossbred pigs, taking advantage of heterosis and breed complementarity (Visscher *et al.* 2000). Crossbred pigs are mostly raised in farms at the commercial level which have lower management and biosecurity conditions compared with nucleus farms. This difference in conditions between commercial and nucleus farms is often reflected in the traits (Dekkers 2007). The same trait when measured in a commercial crossbred animal is not genetically the same as when it is measured in a purebred animal at a nucleus farm. This difference between the traits is reflected in genetic correlations below 1.0, even when the same trait is measured in purebred and crossbred animals. Lutaaya *et al.* (2001) found genetic correlations of 0.62 for growth rate, and 0.32 and 0.70 for backfat thickness between purebred and crossbred phenotypes. Whereas Cecchinato *et al*. (2010) found genetic correlation of 0.25 for piglet survival at birth. A strategy has been proposed in which crossbred animals are used in the training population to subsequently select purebred breeding animals for crossbred performance. This strategy is expected to give a higher response in crossbred performance compared with within-purebred-population selection (Dekkers 2007; Kinghorn *et al.* 2010; Van Grevenhof and Van Der Werf 2015). Besides the increase in response at the crossbred level, using crossbred data in the training population is also appealing because it allows breeding for traits for which phenotypes are scarce in purebreds. Some traits cannot be evaluated in nucleus herds, such as disease traits (Ibañez-Escriche and Gonzalez-Recio 2011).

The strategy of maximizing response to selection of purebreds for crossbred performance by using a crossbred training population has only been evaluated in simulation studies (Dekkers 2007; Kinghorn *et al.* 2010; Van Grevenhof and Van Der Werf 2015). The main issue in performing empirical studies is the need of phenotypes and genotypes of crossbred animals. The collection of these data is costly because this requires, besides genotyping, the individual recording of phenotypes on animals that are kept in group-housing systems and often have no pedigree information. Breeding companies were hesitant to make such investments.

Recently, however, crossbred data for genomic selection in pigs is becoming increasingly important.

In Chapter 5, data on purebred animals were used to predict performance of crossbreds. At the time, the number of genotyped crossbreds was not large enough to be used as a training population. Accuracies of predicting crossbred performance ranged from 0.11 to 0.31 for traits in which the genetic correlation between purebred and crossbred performance ranged from 0.88 to 0.90. These accuracies were not as great as accuracies for within-purebred-population, but they show the predictive ability between purebred and crossbred pigs. For the trait whose accuracy of prediction was zero, a low genetic correlation between purebred and crossbred performance was found (0.31) which is in line with this low accuracy. The predictive ability found for predicting crossbreds with purebred training data indicates that selection in the purebreds will result in a response in the crossbreds when the genetic correlation between purebred and crossbred performance is high.

In Chapter 5, the response variable for genomic prediction was a deregressed breeding value from a routine genetic evaluation. This breeding value was estimated based on records from a mix of purebred and crossbred animals. In practice, there is no problem with the use of a breeding value from a routine genetic evaluation in the evaluation. For research purposes however, it is important to investigate how the choice for purebred, crossbred, or a mix of data used to estimate the breeding values for genomic prediction affects accuracy. In Chapter 6, therefore, we looked into the source of phenotypic information used to estimate the breeding values for the training data set. Training on breeding values of purebred animals estimated using crossbred performance, resulted in more accurate prediction of crossbred genetic merit than training on breeding values of purebred animals estimated using purebred performance; as long as the breeding values that were used as response variable have the same reliability. Likewise, in a simulation study, Esfandyari *et al.* (2015) showed that selecting purebred animals based on crossbred performance data rather than on purebred performance data resulted in a greater response to selection in the performance of crossbred animals.

The results from Chapters 5 and 6 were promising and showed the ability of purebred data to predict performance of crossbred pigs. Thereafter, I wanted to test whether the use of crossbreds in the training population results in greater accuracies than solely using purebreds to select purebreds for crossbred performance. This analysis became possible because more data on crossbred animals became available

(Chapter 7). There was predictive ability when using crossbred phenotypes as training data, however, the accuracies were lower than from using purebred phenotypes. Results of simulation studies (e.g. Dekkers (2007)) that showed greater accuracy from using data on crossbreds rather than on purebred animals in the training population were not confirmed by my results. This discrepancy is explained by the high genetic correlation (>0.90) between purebred and crossbred performance for the traits studied in this thesis. The simulations studies consider a lower genetic correlation between purebred and crossbred performance (0.70 - 0.80) (Dekkers 2007; Van Grevenhof and Van Der Werf 2015). Further studies with other traits with lower genetic correlation between purebred and crossbred performance need to be carried out. I would expect that with lower genetic correlations between purebred and crossbred performance, the benefits from using crossbreds as training population would increase in comparison with using purebreds. With a breeding goal in which all traits have high genetic correlation between purebred and crossbred performance, there would be no need for a crossbred training population, current practice with purebred training would suffice. However, not all traits will have a correlation close to 1, as has been shown by other studies in pigs (Lutaaya *et al.* 2001; Cecchinato *et al.* 2010).

Although greater response to selection is observed in simulation studies from the use of crossbred data for training, these scenarios need to be carefully assessed. Factors such as the reliability of field records and the generation lag could hinder genomic prediction (Ibañez-Escriche and Gonzalez-Recio 2011). As phenotypes will be recorded in crossbreds from commercial farms, the recording system must be well designed and correctly applied because the large number of crossbred animals might be a hindrance to data collection compared with nucleus farms. On top of that, the difference in generations between purebred selection candidates and crossbred pigs, might hamper the genetic gain of genomic selection based on crossbreds. Thus, there is a need for studying whether the additional genetic gains promised by simulations can be confirmed by empirical studies. The additional genetic gains must offset the disadvantages mentioned above.

Using crossbred pigs in the training population to select purebreds for crossbred performance also has an effect on the purebred genetic progress. When genetic correlation between purebred and crossbred performance is high, one will still observe purebred genetic progress. If, however, the genetic correlation is low, one can expect less genetic progress in purebred, or even negative values. With crossbred training populations, the evaluation of breeding program performance will need to shift from analyzing the genetic progress in purebreds to monitoring the improvement of crossbred performance.

2.3.5 Using dominance information for genomic prediction

Dominance is important in crossbreeding schemes as it is the likely basis of heterosis (Xiao *et al.* 1995; Falconer and Mackay 1996; Charlesworth and Willis 2009). Therefore, using a model that accounts for dominance is expected to be beneficial for genomic prediction with a crossbred training population. Hence, I have evaluated genomic prediction when dominance effects are accounted for in the model using a crossbred training population (Chapter 8).

Some studies have reported dominance variance estimates using real pig data and pedigree-based models (Culbertson *et al.* 1998; Norris *et al.* 2010). Estimates of dominance variance are not so precise because they require massive amounts of data especially on full-sib families (Vitezica *et al.* 2013). Dominance variance estimates from pedigree information were found to be zero for gestation length and total number of piglets born (Chapter 8). With genomic information, dominance variance can be estimated more precisely based on heterozygosity of SNP genotypes (Vitezica *et al.* 2013). Studies using genomic data in purebred pigs, showed that non-additive effects are relevant factors contributing to the genetic variation of the studied traits (Su *et al.* 2012; Nishio and Satoh 2014). In addition, they also showed that accounting for the dominance effects improved accuracy of genomic prediction, compared to accounting only for additive effects. Using genomic data from crossbred pigs I showed that, for a trait with dominance variation, accounting for dominance effects can slightly improve genomic predictions compared with accounting only for additive effects (Chapter 8) similar to the reports on purebred pigs mentioned above. Even though there was a slight improvement in prediction from adding the dominance effect, I expect that the inclusion of non-additive effects in routine genetic evaluations is still a long time ahead of us, if breeding companies will ever include them at all. It has been shown that breeding programs should focus on additive effects as they account for more than 50%, and often even 100% of the genetic variation (Hill *et al.* 2008).

Besides a dominance model, a model accounting for breed-specific effects of marker alleles may be relevant in prediction of crossbreeding performance (Ibánez-Escriche et al. 2009). I have found indications that the proportion of genetic variance in crossbred performance differs between the parental purebreds that contributed to the cross (Chapter 8). Such a model, however, needs to be empirically investigated before implementation in breeding programs can be considered.

2.4 Concluding remarks

In the first part of this thesis I describe research that detected genetic markers significantly associated with gestation length, fine-mapped a QTL region for androstenone level, and studied potential pleiotropic effects. I expect that GWAS will continue to be performed because they provide scientifically relevant results, especially with the greater statistical power when more animals will be sequenced or genotyped using HD SNP chips. With more markers, the physical distance between marker and the causative mutation will be shortened, therefore, QTL regions can be fine-mapped. However, finding the causative mutation will require more than just a GWAS using denser genotyping or sequence data. Linkage disequilibrium plays a major role in GWAS and one may require addition functional evidence to distinguish associated variants. The results of GWAS can be incorporated in a MA-GBLUP, to increase the accuracy of genomic prediction compared with GBLUP.

In the second part of this thesis I describe genomic prediction using purebred and crossbred pigs, which is a subject that is highly relevant for pig breeding. Although little has been reported so far, efforts to have more data on crossbred animals have been ongoing and contributed to the analyses performed in this thesis. I have shown that there is predictive ability from using phenotypes of crossbred animals to predict the genetic merit of purebred animals for crossbred performance. Even though the results obtained did not confirm the simulation results, I expect that for other traits with low genetic correlation between purebred and crossbred performance, the simulation results will be confirmed. If confirmed in empirical studies, the use of crossbred training populations for genomic selection will be implemented by breeding companies. The implementation of crossbred training population will, at least in the foreseeable future be without accounting for non-additive effects. Reasons for omitting non-additive effects from prediction models are the large proportion of the total genetic variance explained by additive effects, the increased computational power required to generate for example a genomic dominance matrix, and the negligible added-value to accuracy shown so far from adding dominance to genomic prediction.

References

Animal QTLdb. http://www.animalgenome.org/QTLdb. Accessed 21 Aug 2015

Calus MPL, Meuwissen THE, De Roos APW, Veerkamp RF (2008) Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–61.

Cecchinato A, de los Campos G, Gianola D, et al (2010) The relevance of purebred information for predicting genetic merit of survival at birth of crossbred piglets. J Anim Sci 88:481–490.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nat Rev Genet 10:783–96.

Chen L, Vinsky M, Li C (2015) Accuracy of predicting genomic breeding values for carcass merit traits in Angus and Charolais beef cattle. Anim Genet 46:55–59.

Culbertson MS, Mabry JW, Misztal I, et al (1998) Estimation of dominance variance in purebred Yorkshire swine. J Anim Sci 76:448–451.

Daetwyler HD, Capitan A, Pausch H, et al (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet 46:858–865.

De Roos APW, Hayes BJ, Goddard ME (2009) Reliability of genomic predictions across multiple populations. Genetics 183:1545–53.

Dekkers JCM (2007) Marker-assisted selection for commercial crossbred performance. J Anim Sci 85:2104–14.

Duijvesteijn N, Knol EF, Merks JWM, et al (2010) A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genet 11:42.

Esfandyari H, Sørensen AC, Bijma P (2015) Maximizing crossbred performance through purebred genomic selection. Genet Sel Evol 47:1–16.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, 4th edn. Longman, Harlow, England

Fujii J, Otsu K, Zorzato F, et al (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science 253:448–451.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml user guide release 3.0.

Grindflek E, Lien S, Hamland H, et al (2011) Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. BMC Genomics 12:362.

Harris BL, Johnson DL, Spelman RJ (2008) Genomic selection in New Zealand and the implications for national genetic evaluation. Proc. Interbull Meet. Niagara Falls, p 325–330

Hayes BJ, Bowman PJ, Chamberlain AC, et al (2009a) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol 41:51.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009b) Invited review: Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–43.

Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4:e1000008.

Ibánez-Escriche N, Fernando RL, Toosi A, Dekkers JCM (2009) Genomic selection of purebreds for crossbred performance. Genet Sel Evol 41:12.

Ibáñez-Escriche N, Forni S, Noguera JL, Varona L (2014) Genomic information in pig breeding: Science meets industry needs. Livest Sci 166:94–100.

Ibañez-Escriche N, Gonzalez-Recio O (2011) Review. Promises, pitfalls and challenges of genomic selection in breeding programs. Spanish J Agric Res 9:404–413.

Kinghorn BP, Hickey JM, Werf JHJ Van Der (2010) Reciprocal recurrent genomic selection for total genetic merit in crossbred individuals. Proc. 9th WCGALP. p 36

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756.

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. J Dairy Sci 92:4656–4663.

Lund MS, Su G, Janss L, et al (2014) Genomic evaluation of cattle in a multi-breed context. Livest Sci 166:101–110.

Lutaaya E, Misztal I, Mabry JW, et al (2001) Genetic parameter estimates from joint evaluation of purebreds and crossbreds in swine using the crossbred model. J Anim Sci 79:3002–7.

Merks JWM, De Vries AG (2002) New sources of information in pig breeding. Proc. 7th WCGALP. p 3–10

Meuwissen T, Goddard M (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–631.

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–29.

Misztal I (2011) FAQ for genomic selection. J Anim Breed Genet 128:245–246.

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92:4648–4655.

Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet 124:342–355.

Nishio M, Satoh M (2014) Including dominance effects in the genomic BLUP method for genomic evaluation. PLoS One 9:e85792.

Norris D, Varona L, Ngambi JW, et al (2010) Estimation of the additive and dominance variances in SA Duroc pigs. Livest Sci 131:144–147.

Purfield DC, Bradley DG, Evans RD, et al (2015) Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. Genet Sel Evol 47:47.

Ramos AM, Crooijmans RPMA, Affara NA, et al (2009) Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One 4:e6524.

Sahana G, Guldbrandtsen B, Thomsen B, et al (2014) Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. J Dairy Sci 97:7258–7275.

Su G, Christensen OF, Ostersen T, et al (2012) Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One 7:e45293.

Tiezzi F, Maltecca C (2015) Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genet Sel Evol 47:1–13.

Tusell L, Pérez-Rodriguez P, Forni S, et al (2013) Genome-enabled methods for predicting litter size in pigs : a comparison. Animal 7:1739–1749.

Van Eenennaam AL, Weigel KA, Young AE, et al (2014) Applied Animal Genomics: Results from the Field. Annu Rev Anim Biosci 2:105–139.

Van Grevenhof IEM, Van Der Werf JHJ (2015) Design of reference populations for genomic selection in crossbreeding programs. Genet Sel Evol 47:1–9.

VanRaden PM, Van Tassell CP, Wiggans GR, et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24.

Veroneze R (2015) Linkage disequilibrium and genomic selection in pigs. PhD Thesis. Wageningen University

Visscher P, Pong-Wong R, Whittemore C, Haley C (2000) Impact of biotechnology on (cross)breeding programmes in pigs. Livest Prod Sci 65:57–70.

Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet 9:255–266.

Vitezica ZG, Varona L, Legarra A (2013) On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195:1223–1230.

Wu X, Lund MS, Sahana G, et al (2015) Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. Genet Sel Evol 47:50.

Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754.

Zhang Z, Liu J, Ding X, et al (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. PLoS One 5:1–8.

Summary

Summary

In the last decade, high-throughput genomic information became available for most livestock species. Efficient use of this information is important for the competitiveness of a breeding company. Application of genomic selection (GS) in pigs, may require different strategies from what is currently applied in dairy cattle because the end product in pig production is a crossbred animal. In this work, I explored the use of genomic information for the genetic improvement of purebred and crossbred pigs. Firstly, working mainly in purebred animals, regions affecting gestation length (Chapter 2) and androstenone level (Chapter 3) were detected in the pig genome by genome-wide association and fine-mapping. Also, potential pleiotropic effects of the androstenone level quantitative trait locus (QTL) on reproductive traits were studied (Chapter 4). Secondly, we investigated the potential of GS in pig breeding by determining the accuracy of genomic prediction using different strategies. These strategies varied in training and validation populations, selected from multiple purebred lines and their crossbred offspring, different data types and models.

Genome-wide association study (GWAS) identified two QTL regions for gestation length, one in the Dutch Landrace and one in the Large White (Chapter 2). Three associated SNP were detected in a QTL region spanning 0.52 Mbp on *Sus scrofa* chromosome (SSC) 2 in Dutch Landrace and for the Large White, four associated SNP were detected in a region of 0.14 Mbp on SSC5. The region of a previously detected QTL for androstenone level on SSC6 was fine-mapped, narrowing the region down from 3.75 Mbp to 1.94 Mbp and identifying a candidate mutation in *SULT2A1* (Chapter 3). This fine-mapped region was evaluated for possible pleiotropic effects on production and reproduction traits in pigs (Chapter 4). No unfavorable pleiotropic effects were found, indicating that using the studied marker for selection would not unfavorably affect the other relevant traits.

In the later chapters I have investigated the potential of different strategies for the implementation of GS in pig breeding when the aim is to improve crossbred performance. Within-population prediction was showed considerably high accuracy of prediction (Chapters 5 and 6) while across-population prediction, evaluated in Chapter 5 had accuracies close to zero. Multi-population prediction, where combinations of Dutch Landrace and Large White animals plus their cross were used as training showed that adding data from other populations did not improve the accuracy except when predicting the F1 cross with records from both parental populations added to the F1 training data. When only purebred data was used, there was some predictive ability for crossbred performance (Chapter 5). In the first study the training data contained a mix of records measured on purebred and crossbred animals. In Chapter 6, therefore, the source of training data was clearly separated into purebred and crossbred records. Training on breeding values of purebred animals that were estimated using crossbred offspring performance, resulted in more accurate prediction of their crossbred genetic merit compared with training on breeding values of those same animals, estimated using purebred offspring performance. Genotyped and phenotyped crossbreds in the training population were expected to have higher accuracies when predicting genetic merit for crossbred performance. However, in Chapters 5 and 6 we did not test this strategy because sufficient genotyped crossbred were lacking at that time. Later, with more crossbred data, we evaluated this strategy and the accuracies were not improved over the use of genotyped and phenotyped purebreds (Chapter 7) mainly due to the high genetic correlation between purebred and crossbred performance for the studied traits. Finally, the inclusion of dominance in the model, with a crossbred training population was evaluated. For a trait that had dominance variation, accounting for dominance effects can be slightly beneficial for genomic prediction compared with a model that accounts only for additive effects.

Finally, in Chapter 9, the relevance of the findings was discussed, how breeders can benefit from the combination of genomic selection with the information of individual QTL. To finalize, I make suggestions for future studies and how breeders can make use of the results generated in the thesis.

Acknowledgements

Acknowledgements

This PhD degree would not be possible without the help and contribution of many people. I am very grateful to everybody that directly or indirectly contributed to the completion of this thesis.

Thanks to my three supervisors that provided me knowledge, confidence and friendship to finalize this PhD degree.

DJ, thanks for everything, especially during the time I spent in Uppsala. It has always been nice to chat with you in any occasion, always with nice jokes and emoticons via Skype ☺. Even though apparently you were not that happy when I was singing while getting to my office early in the morning… "Too easy this PhD life" hehe. Now instead of keeping the piggies always pinky I will try to keep the cows always spotted ☺.

John, besides the technical genetic part, be sure that I learnt a lot on how to look at a manuscript in a more critical editorial way because of you. Your comments and modifications are always useful and I really look up to your writing style. At the same time I learnt a lot during your daily supervision that I will take and apply to my researcher life. I hope you didn't get annoyed by the many times I asked whether you had revised my manuscript. Thanks for the support given to me during these four years regarding trip to the US, writing more papers, job applications, having holidays, everything. You always gave me the support to believe that was possible and indeed it was…

Martien, thank you for the support even when I changed the direction of the PhD project deviating from your main field of expertise. You are very busy yet your door was always open for me whenever I had doubts or comments. Thanks for being my promoter in such a unique opportunity that I had of doing my PhD abroad.
Thanks to my committee members (opponents) for reading the thesis and hopefully we will have a fruitful discussion during the defense ceremony.

I very much appreciate the EGS-ABG initiative including all coordinators, secretaries and participants of this amazing program. Thanks to the European Union and Topigs Norsvin for funding my 4 years of PhD studies.

Thanks to Wageningen University and the Swedish University of Agricultural Sciences for the opportunity to study in such renowned universities.

Thanks to Ada, Lisette and Monique from ABGC as well as to Helena, Harriet and Monica from the Animal Breeding and Genetics department at SLU for all the help with paperwork.

To all staff and people that I met from WUR and SLU, I learnt a lot during QDGs, seminar series, TLMs, genomics meetings, coffee- and lunch-breaks, lab meetings, etc.

Thanks to Topigs people for being always very open to receive me in the company, to provide data and to discuss results, doubts, etc. You are part of a nice company and wish you all the success! I really appreciated the way PhD candidates are treated there, taking part in meetings, colloquium with the doors always open to receive us.

Thanks Jack Dekkers for receiving me in Ames as a visiting scholar, also Jian and Rohan. It was a great time spent in Ames learning a lot, getting to know very nice people and facing a very nice cold weather! Thanks to the whole Animal Breeding and Genetics group in Ames for receiving me very well during the 3 months! Valeu Nicola pela ajuda também, tocar um violão, cantar um Raimundos, torcer pro São Paulo e ter conhecido a sua sensacional família!

Obrigado a todo o pessoal que convivi como amigo ou estudante durante meus estudos na Universidade Estadual de Maringá e da Universidade Federal de Viçosa, vocês sempre me apoiaram e me deram forças! Um agradecimento especial para o Elias que despertou meu interesse pelo melhoramento! Também agradeço o Paulo Sávio, Simone e o Fabyano por todo o período que estive em Viçosa e após também, além da ótima relação de amizade que tenho com vocês!

Muito obrigado ao Brenno Harry Oliveira Silva por desenvolver a criação e design gráfico da capa! Obrigado também ao Cleibe Pinto por ter supervisionado todo esse processo!

These 4 years of PhD life passed by in a blink of an eye. One of the main reasons for that was the amazing group of people that I had the opportunity to know and to build up very nice friendship. There are too many awesome people that I have met in Wageningen and Uppsala, as well as in conferences, EGS-ABG, courses, etc. Thanks for everything, I really enjoyed spending plenty of time discussing scientific matters, the PhD life, having barbecues, kebabs, ribs, Chinese restaurant, playing games, pub-quizing and the list goes on.

Special thanks to a bunch of people, Nancy, words are inadequate to express my gratitude of having your friendship! Hehehe! We really had a nice time during these 4 years in both Wageningen and Uppsala. Thanks for all the conversations that we had about life, PhD, papers, jokes, etc. Thank you very much for you English lessons and singing performances! I wish you all the best! Sandrine muito obrigado pela sua amizade, tanto na Holanda como na Suécia, apesar da sua personalidade forte que eu sempre soube contorná-la com uma boa risada, você sempre quer o bem das pessoas e sei que sempre posso contar com você, assim como você pode contar comigo! Katrijn thanks for all the talks, parties, rides, singing, sewing, trips, dinners, house renting, and especially your friendship! Gus, Mr. Munni, thanks for a true friendship. You really helped me to mingle and to develop my social skills. All the dinners at your place, the amazing hamburger with cheese inside! Johnny Cash songs! Australian day! All the room sharing during conferences and drinking from the bottle sideways! All the Aussie lessons! Get a dog up ya! Mathieu thanks for all the fish and French lessons, nice talks, songs, trips and for an OTH friendship! Naomi, thanks a lot for your friendship as well as a lot of Guinness, M&Ms, Dutch lessons that I will have with your daughters now that we are neighbours hehe! Juanma, the Spanish guy! Thanks for the friendship, always sharing some good stories, good food and losing football matches on the pro-nights hehe! ☺. Claudia valeu pela amizade, por sempre ter assunto bom pra conversar, um ótimo humor e pelo seu amor pelo Brasil heheheheh! Obrigado Renata e Lucas por dividirem um bom tempo por aqui na Holanda, revivendo os tempos de Viçosa hehe! Thanks also to my office-mates Dianne, Marzieh, Yvonne, always sharing some food, jokes and scientific questions! Thanks to the “PhD mafia” in Uppsala, Merina, Thu, Agnese, Berihu, Ahmed, Chrissy, Bingjie, Josh, Alberto, Sangeet, Fabiana, Iris, Axel, Jonas! Some nice table-tennis matches, charades, a lot of singing and of course good food! Each and every one of you really made every weekend fun in Uppsala!

Marcola, pensa num cara parceiro e que sempre estava presente pro que der e vier nesses 4 anos por aqui! Ainda me lembro de quando você veio pra Wageningen para nos encontrarmos de novo aqui na Holanda e eu comi o croquete com mostarda hehehe! Todas as viagens que fizemos e ainda faremos, discussões científicas e não científicas, academia, ótimas conversas e refeições, ou seja, qualquer coisa. Não tenho muitas palavras, apenas um obrigado por ser seu amigo!

Gabriel, big G! Since the first days in Wageningen you were always with a smile on your face and ready to share a nice conversation. I really enjoy spending time with you! Even though you were in Denmark for 2 years whenever we would meet it seemed that we had seen each other just a little while ago... Thanks a lot for your friendship and be sure que esta amizade é sincera!

Família, não tem nada mais importante do que isso! Muitíssimo obrigado por todo o carinho, ajuda, suporte, conversas, visitas durante não só esse período do doutorado, mas durante minha vida toda! Muito obrigado por me permitirem ir atrás das minhas conquistas sempre me apoiando! Eu realmente sou muito sortudo de ter uma família tão boa!

Davi, obrigado por sempre querer falar no Skype já que as vezes não sou muito adepto da comunicação hehe! Bruna, também muito obrigado! Tia Ete, obrigado pela força e por sempre estar interessada em como eu estou e com as coisas vão por aqui! Mami e Papi, realmente não tenho palavras para vocês, apenas de que tudo que sou é fruto do que aprendi com vocês! Estou julgando que sou algo bom hehe! De verdade, muito obrigado por tudo!

Vica, os 4 anos se passaram! Ainda lembro de quando me disse que valia a pena tentar! E valeu! Obrigado pelo amor, amizade, sinceridade, tudo que você me proporciona! Agora começando uma nova empreitada nas nossas vidas juntos! Muito obrigado também a sua família pelo apoio que sempre me(nos) dão! ☺