Genomics of heterosis and egg production in White Leghorns

Esinam Nancy Amuzu-Aweh

WAGENINGEN UR
For quality of life

SLU

ACTA UNIVERSITATIS AGRICULTURAE SUECIAE DOCTORAL
THESIS No. 2020:6

Genomics of heterosis and egg production in White Leghorns

Esinam Nancy Amuzu-Aweh

Thesis committee

Promotor
Prof. Dr H. Bovenhuis
Personal chair at Animal Breeding and Genomics
Wageningen University & Research

Co-promotors
Dr P. Bijma
Assistant professor, Animal Breeding and Genomics
Wageningen University & Research

Prof. Dr D. J. de Koning
Professor of Animal Breeding
Swedish University of Agricultural Sciences

Other members (Assessment committee)
Prof. Dr R. G. F. Visser, Wageningen University & Research
Dr A. Wallenbeck, Swedish University of Agricultural Sciences, Uppsala, Sweden
Prof. Dr S. E. Aggrey, University of Georgia, USA
Prof. Dr J. Bennewitz, University of Hohenheim, Stuttgart, Germany

The research was conducted under the joint auspices of the Swedish University of Agricultural Sciences, Sweden, and the Graduate School of Wageningen Institute of Animal Sciences, The Netherlands, and as part of the Erasmus Mundus Joint Doctorate Program “EGS-ABG”.

Genomics of heterosis and egg production in White Leghorns

Esinam Nancy Amuzu-Aweh

ACTA UNIVERSITATIS AGRICULTURAE SUECIAE DOCTORAL THESIS No. 2020:6

Thesis

submitted in fulfillment of the requirements for the joint 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 Friday 6 March 2020
at 4 p.m. in the Aula of Wageningen University.

Amuzu-Aweh, E.N.
Genomics of heterosis and egg production in White Leghorns
194 pages.

Joint PhD thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden and Wageningen University, Wageningen, the Netherlands (2020)
With references, and with summaries in English and Swedish.

ISSN: 1652-6880
ISBN (print version): 978-91-7760-530-0
ISBN (electronic version): 978-91-7760-531-7
ISBN: 978-94-6395-239-2
DOI: 10.18174/508397

Abstract

Amuzu-Aweh, E.N. (2020). Genomics of heterosis and egg production in White Leghorns. Joint PhD thesis, between Swedish University of Agricultural Sciences, Sweden and Wageningen University, the Netherlands

Crossbreeding is practiced extensively in commercial breeding programs of many plant and animal species, in order to exploit heterosis, breed complementarity, and to protect pure line genetic material. The success of commercial crossbreeding schemes depends on identifying and using the right combination of breeds, lines or varieties that produce the desired crossbred offspring. Currently, the selection of pure lines is based on the results of "field tests", during which the performance of their crossbreds is assessed under typical commercial settings. Field tests are time-consuming, and also constitute a large percent of the costs of commercial crossbreeding programs. The research in this thesis therefore set out mainly to develop models for the accurate prediction of heterosis in White Leghorn crossbreds, using genomic information from their parental pure lines. Predicted heterosis could be used as pre-selection criteria, thus substantially reducing the number of crosses that need to be field-tested. In **Chapter 1**, I give an overview of the history of selective breeding in laying hens, and introduce heterosis and its genetic basis. In **Chapter 2**, based on a dominance model, we showed that a genome-wide squared difference in allele frequency between parental pure lines (SDAF) predicts heterosis in egg number (EN) and egg weight (EW) at the line level with an accuracy of ~0.5. With this accuracy, one can reduce the number of field tests by 50%, with only ~4 loss in realised heterosis. In laying hens, selection pressure is highest on the sires. We therefore went further to develop a model to predict heterosis at the individual sire level, in order to exploit the variation between sires from the same line. We found that the within-line variation between sires in our data was very small (0.7% of the variation in predicted heterosis), and most of the variation was explained by across-line differences (90%) **(Chapter 3)**. Quantitative genetic theory shows that heterosis is proportional to SDAF and the dominance effect at a locus. In **Chapter 4,** we estimated variance components and dominance effects of single nucleotide polymorphisms (SNPs) on EN and EW in White Leghorn pure lines. We found that dominance variance accounted for up to 37% of the genetic variance in EN, and up to 4% of that in EW. We then used the estimated dominance effects to calculate dominance-weighted SDAFs for EN and EW between parental pure lines, and showed that prediction of heterosis based on a weighted SDAF would yield considerably different ranking of crosses for each trait, compared with a prediction based on the raw SDAF. This implies that different crosses would

be selected depending on the criterion used to predict heterosis. To gain an insight into the genetic architecture of EN and EW, in **Chapter 5** we performed genome-wide association studies using data on 16 commercial crossbred populations. We did not identify any significant SNPs for EN, indicating that EN is a highly polygenic trait with no large quantitative trait loci segregating in the populations studied. For EW, however, we identified several significant SNPs. One explanation for these results is that EN has been under intense directional selection for several decades, whereas EW has been under less-intense, stabilising selection. Finally, in the general discussion of this thesis **(Chapter 6)**, I discuss the genomic prediction of heterosis, focusing on possible reasons for the lack of a consensus on the approach to predict heterosis, even after decades of research. I also discuss new opportunities for the genomic prediction of heterosis, considering the advancements in genotyping and computation methods. Lastly, I give an example of the application of results from this thesis in crossbreeding programs.

For my family

Contents

CHAPTER 1

General Introduction

1.1 Introduction

Chickens provide 92% of all eggs consumed globally, and most of this comes from commercial breeding flocks (FAO, 2018). Over the years, selective breeding for improved genetic value of chickens, and the use of crossbreeding schemes, have made it possible for laying-hen industries to meet the ever-rising demand for good quality eggs. In recent times, animal breeders are interested in developing methods to further utilise genomic information of selection candidates in order to increase the efficiency of breeding programs.

This thesis is about the use of genomic information to optimise commercial crossbreeding schemes in laying hens. As an introduction to the topic, first I will give an overview of selective breeding in laying hens – its history, the use of crossbreeding, and the evolution of breeding goals. Next I will describe heterosis, which is one of the main benefits of crossbreeding, and is the focal point of my thesis research. I will then end with a section on the motivation, objectives and outline of this thesis.

1.2 Selective breeding in laying hens

1.2.1 History

Present-day domestic fowls, *Gallus gallus domesticus,* are descendants of the red jungle fowl, *Gallus gallus* (Crawford, 1990), and are also believed to have some genetic contribution from the grey jungle fowl, *Gallus sonneratii* (Eriksson et al., 2008). The exact time and place of the domestication of chickens remains unclear, but it was probably in South East Asia at about 6000 BC. One thing is for certain though – chickens were 'domesticated' and spread to Europe and America for their participation in cock fighting – not for food (Crawford, 1990; Thomson, 1964; Yamada, 1988). It was the Romans who first began to view chickens as a source of food, and started developing their potential for agriculture (Thomson, 1964).

Most of the commercialisation of layer breeding in Europe and North America began in the early 20th century. Around that same time, production moved from the backyard system to an intensive production system (Elson, 2011). Next began the development of specialised production units, and with it, the need for advanced genetic programs. Therefore, from the 1950's up until the year 2000, pedigree information, selection indices and best linear unbiased prediction (BLUP) breeding values were used as selection criteria (Arthur and Albers, 2003); prior to this, breeders had been practicing selection on own phenotype for females and progeny testing for males. In addition to the other advancements in genetic programs,

chicken breeders started to develop specialised 'pure' lines, and began using crossbreeding schemes to produce the commercial flocks.

Crossbred layers were highly productive, and therefore the success of crossbreeding resulted in the merger of smaller breeding firms to form fewer but larger breeding companies that had the resources to carry out the intensive selection programs required to develop specialised pure lines, and could produce large numbers of commercial crossbred day-old chicks for sale. Important factors that made large-scale production of day-old chicks possible were: 1) the use of artificial insemination, which allowed flexibility in mating ratios and efficient propagation of superior genetics; 2) the development of large-scale artificial incubators which made it possible to hatch hundreds of thousands of chicks simultaneously; and 3) the use of artificial lighting systems which influenced laying behaviour, thereby enabling year-round lay. All these advancements in the industry came hand-in-hand with improvements in sanitation, disease control and vaccination.

In 2001, genomic selection (GS), where animals are selected based on genomic breeding values estimated from genome-wide marker effects, was introduced (Meuwissen et al., 2001). A few years later, GS started being applied in experimental flocks, and by 2013, it had been applied to a commercial flock (Wolc et al., 2016). Genomic selection currently forms part of the routine evaluation in commercial laying-hen breeding programs, and has resulted in substantial increases in the accuracy of selection and genetic gain.

1.2.2 Crossbreeding

Crossbreeding is the mating of individuals from different breeds (or lines/varieties/ strains) with the aim of producing offspring that have a combination of the desired characteristics of both parental breeds and perform better than their parents. Deliberate and organised crossbreeding is believed to have begun in maize (Bennetzen and Hake, 2009), and following that, breeding programs for several plants, *e.g.* wheat, rice, tomato, sorghum and some oilseeds, developed inbred lines and produced crossbreds (hybrids) as well. Learning from this, crossbreeding also started extensively in laying hens, to produce egg-layers that are either three- or four-way crossbreds. Crossbreeding is also practiced in the commercial breeding programs of other animal species, *e.g.* pigs, beef cattle, sheep and goats.

Laying-hen breeding companies usually maintain multiple 'pure' lines and therefore one company may produce several types of commercial crossbreds. The best

combination of pure lines to be used in each cross was, and still is, partly determined by performing field-tests during which many pure line combinations are made, and the performance of their crossbred offspring is evaluated for several traits. Crossbred performance is then used to make informed decisions on which pure lines to cross to produce the best commercial crossbred flocks.

A widely used breeding structure for laying hens is in the form of a pyramid (Figure 1.1). At the top of the pyramid are nucleus flocks made up of pure lines. The nucleus is where intense selection pressure is applied, and thus where genetic progress is made. Breeders usually focus on improving specific traits in each pure line, or developing pure lines that are suited for specific production systems and environments. In addition, most pure lines are specialised as either sire or dam lines. The next level of the pyramid is the multiplying unit, with the function of increasing the number of purebred individuals. It is also referred to as the great-grand-parent level. After this comes the level with the grand-parents of the commercial flock, followed by a level where the parents (sires and dams) of the commercial flock are. The parent level is the first level that has crossbreds: either both the sires and dams are products of a two-way cross, *i.e*. they are products of a pure line × pure line mating, or only the dams are two-way crossbreds and the sires are purebreds. The next and final level of the pyramid is made up of the commercial flock. Depending on which parents were used, the birds here are either three-way or four-way crossbreds.

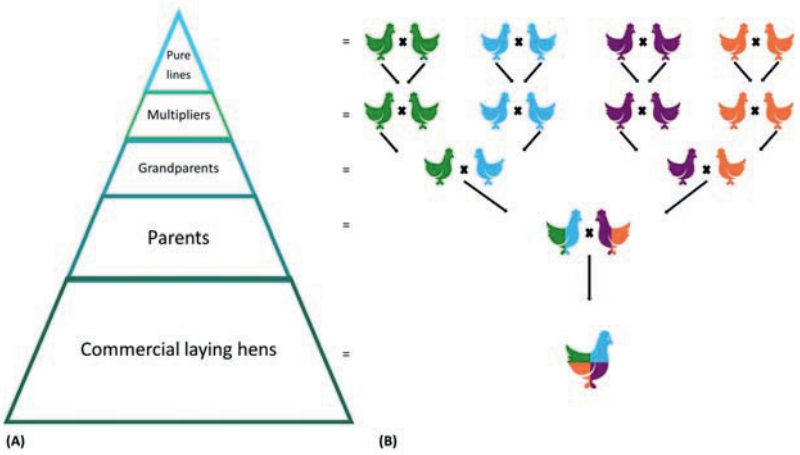

Figure 1.1 Breeding structure used for commercial laying hens. (A) Pyramid breeding structure (B) Four-way terminal crossbreeding scheme.

Crossbreeding has been successful in laying hens for a number of reasons: 1) the exploitation of heterosis in crossbred individuals; 2) it allows breeding companies to protect their genetic material, since it is not beneficial for farmers to use the commercial crossbreds for breeding purposes; 3) it makes sexing of day-old chicks quite straightforward - *e.g.* using the sex-linked gene (*K* = slow feathering; dominant and *k* = fast feathering; recessive). If *k* is fixed in the sire lines (Z^k/Z^k) and *K* in the dam lines (Z^K/W), then crossing these lines will produce males that are all slow feathering (Z^K/Z^k), and females that are all fast-feathering (Z^k/W); 4) the benefit of breed complementarity, *i.e.* sire and dam lines can be selected for different traits, such that they complement each other. For example, in the sire lines, more emphasis is placed on traits like feathering, behaviour, feed efficiency, egg size and liveability while in the dam lines great focus is placed on egg production, egg quality and liveability. This results in a commercial crossbred that has an ideal combination of all these traits.

1.2.3 Breeding goals

From the onset of commercial breeding up until the year 2013, selection pressure was mainly on productivity (Neeteson-van Nieuwenhoven et al., 2013). One can conclude that in that respect, breeders have been successful – both for the breeder hens, where from the 1980's to 2010, there has been an increase of 15 - 20 in the number of day-old chicks produced by one breeding hen per year (Van Sambeek, 2011), and for the commercial layers, where the average number of eggs laid /hen/year increased from 190 in 1950 to 309 in 1998 (Albers, 1998). In 2011, Van Sambeek reported that the genetic progress in commercial hens was equivalent to 2.5 additional eggs/hen/year (Van Sambeek, 2011).

Breeding goals change over time, however, in response to new knowledge on the biological background of traits, consumer preferences, the production environment, awareness of the importance of the health and welfare of animals, food quality and safety, and the impact of animal production on the environment. For example, the ban on using conventional battery cages in the European Union (Council Directive 1999/74/EC) and on beak trimming in several countries, made traits like feather pecking, cannibalistic behaviour, the ability to produce in free-range or floor systems, and good nesting behaviour more important (Muir et al., 2014). Welfare issues related to induced moulting of commercial laying hens have also led to breeding goals geared towards increasing persistency of lay – to produce a hen that lays 500 eggs in an extended laying cycle of 100 weeks, without moulting (Van Sambeek, 2011). As a result of all these changes, current breeding objectives are made up of a selection index that includes several traits. Productivity is still an

important trait, but more the efficiency of production rather than the level of production.

In summary, the main milestones that led to the development of modern-day selective breeding in commercial laying hens are (not necessarily in this order):

- formation of specialised sire and dam lines
- the effective use of crossbreeding schemes to exploit heterosis and protect genetic material
- advances in reproductive technologies: artificial insemination, incubation and hatching, lighting programs/technologies to influence laying behaviour
- improvement in criteria for selecting animals, through the application of quantitative genetics theory, statistics, and BLUP breeding values
- availability of genomic markers and genomic selection methodology to increase the accuracy of selection and genetic gain

With the current level of experience, increasing knowledge of genetics, genomic selection, improved housing, management and disease control, there is still a lot of potential to develop the laying-hen industry even further.

1.3 Heterosis

Heterosis or hybrid vigour is the superiority of a crossbred individual compared with the average of its purebred parents (Dobzhansky, 1950; Shull, 1952, 1914), and is the main benefit of crossbreeding (Fairfull, 1990). In plants, where fully inbred lines are used to produce the crossbreds, heterosis is generally higher than in animals, where the 'pure' lines that produce the crossbreds are not deliberately inbred.

Yield advantage of crossbred over purebred maize ranges from about 10% to as much as 72% (summarised in Hallauer and Miranda, 1988). In animals, a wide range of heterosis percentages are found in literature: -3 to 40% in laying hens (Fairfull, 1990), -4 to 38% in beef cattle (Gosey, 2005; Kress and Nelsen, 1988) and 2 to 18% in sheep (Nitter, 1978). The general trend in animals is that heterosis is more pronounced in traits that have a low heritability, *e.g*. fertility, disease resistance and longevity – than in traits with relatively high heritability like growth and egg number.

1.3.1 Genetic basis of heterosis

No consensus has been reached on the genetic basis of heterosis; what can be agreed upon is that it is complex, trait-specific and approximately proportional to

the difference in allele frequency between the parental populations (Falconer and Mackay, 1996). Three hypotheses are generally proposed as possible explanations for the genetic mechanisms underlying heterosis: 1) the dominance hypothesis is based on the observation that most deleterious alleles are recessive, and thus attributes heterosis to the masking of these deleterious recessive alleles from one parental line by dominant alleles in the other parental line; 2) the overdominance hypothesis attributes heterosis to advantageous combinations of alleles at heterozygous loci, thereby making the heterozygote superior to either homozygote; and 3) the epistasis hypothesis assumes that interactions among loci lead to heterosis(Crow, 1999; Goodnight, 1999; Lamkey and Edwards, 1999; Lynch and Walsh, 1998). Related to both the dominance and overdominance hypotheses, quantitative genetic theory predicts the presence of heterosis when there is directional dominance. If some loci have positive dominance and others have negative dominance, their effects can cancel out. Directional dominance occurs when $\bar{d} \neq 0$. With directional dominance, heterosis is proportional to the squared difference in allele frequency between parental pure line populations:

$$\boldsymbol{Heterosis} = (\boldsymbol{p_i} - \boldsymbol{p_j})^2 \boldsymbol{d} \qquad \text{Eq. 1.0}$$

where p_i and p_j are the allele frequencies at a particular locus in parental populations *i* and *j* respectively, and *d* is the dominance deviation at that same locus (Falconer and Mackay, 1996). This means that if the two populations do not differ in allele frequency, and/or there is no directional dominance, heterosis will not be observed. Equation 1.0 is the basis of my thesis research.

1.4 This thesis

1.4.1 Motivation

The success of commercial crossbreeding schemes depends on identifying and using the right combination of breeds, lines or varieties that will produce offspring that fit customers' requirements. The focus of my PhD thesis is on situations where multiple pure lines are available to produce multiple crossbred products, as is typical in commercial laying-hen breeding companies. As mentioned earlier, crossbreeding schemes for laying hens – as well as other plant and animal species – use results from field tests in order to identify the best combinations of pure lines to use to produce the commercial crossbreds. These field tests are time-consuming, labour-intensive and expensive, and as the number of parental pure lines increases, it becomes less feasible to field-test all possible combinations of pure lines. Crossbreeding schemes would therefore be more efficient if crossbred performance could be predicted

based on purebred information, because one would know beforehand which combinations of pure lines would give the best crossbred offspring.

The mean phenotypic value of a cross can be partitioned into pure line averages and heterosis. The pure line average can be inferred from the phenotype of the purebred individuals, however, the heterosis component cannot. For this reason, the prediction of heterosis has been of interest to scientists for decades. Quantitative genetic theory shows that when heterosis is due to directional dominance, heterosis is proportional to the squared difference in allele frequency between parental pure lines (Falconer and Mackay, 1996). Stemming from this, several past studies used genetic markers to calculate numeric measures of the divergence between populations, *e.g.* modified Rogers' distance (Wright, 1984) and Nei's genetic distance (Nei, 1972), and estimated correlations between these variables and crossbred performance or heterosis. Results were inconclusive – both in plants and animals – and the general agreement was that a higher number of molecular markers with genome-wide coverage would be needed for further studies (Atzmon et al., 2002; Balestre et al., 2009; Gavora et al., 1996; Haberfeld et al., 1996; Minvielle et al., 2000; Reif et al., 2003 and reviews by Dias et al., 2004; Krishnan et al., 2013).

The current availability of genomic data gives the opportunity to revisit the prediction of heterosis by providing a large number of genome-wide markers and also the opportunity to explore the estimation of non-additive effects. It is therefore of interest to investigate the possibilities to predict heterosis using a large number of genomic markers, and this thesis research is the first to do so for laying hens.

1.4.2 Objective and thesis outline

The main objective of this thesis was to optimise the use of genomic information in commercial crossbreeding schemes of laying hens by developing methods for the prediction of heterosis. We also expected to gain insight on the genetic mechanisms behind heterosis, and to identify genomic regions associated with traits of economic importance. In **Chapter 2**, we investigated whether differences in frequencies of single nucleotide polymorphism (SNP) alleles between parental pure lines was predictive of heterosis at the population level. In **Chapter 3**, we investigated whether individual sire genotypes could be used to predict heterosis at the individual level, in order to exploit the variation between sires from the same pure line, and further increase realised heterosis in crossbred offspring. Since directional dominance is necessary for heterosis to be expressed, in **Chapter 4**, first we estimated dominance

variance and SNP effects for egg number and egg weight, and then discussed the possibility of predicting heterosis by weighting SNPs by their estimated dominance effects. In **Chapter 5**, we explored the genetic architecture of egg number and egg weight in crossbred laying hens by performing a genome-wide association study. Finally, in **Chapter 6**, the General Discussion, I summarise the findings from my research, then discuss the genomic prediction of heterosis, focusing on possible reasons for the lack of a consensus on an approach to accurately predict heterosis. I also discuss opportunities for the genomic prediction of heterosis, considering the advancements in genotyping and computation methods. Next, I give an example of the application of results from this thesis in crossbreeding programs.

1.5 References

Albers, G.A.A., 1998. Future trends in poultry breeding. World Poult. 14, 42–43.

Arthur, J.A., Albers, G.A.A., 2003. Industrial perspective on problems and issues associated with poultry breeding. Poult. Genet. Breed. Biotechnol. 1, 12.

Atzmon, G., Cassuto, D., Lavi, U., Cahaner, U., Zeitlin, G., Hillel, J., 2002. DNA markers and crossbreeding scheme as means to select sires for heterosis in egg production of chickens. Anim.Genet. 33, 132–139.

Balestre, M., Von Pinho, R.G., Souza, J.C., Oliveira, R.I., 2009. Potential use of molecular markers for prediction of genotypic values in hybrid maize performance. Genet Mol Res 8, 1292–1306.

Bennetzen, J.L., Hake, S.C., 2009. Handbook of maize: genetics and genomics. Springer.

Crawford, R.D., 1990. Origin and history of poultry species. Poult. Breed. Genet. 1–41.

Crow, J.F., 1999. Dominance and Overdominance, in: Coors, J.G., Pandey, S. (Eds.), The Genetics and Exploitation of Heterosis in Crops. American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., Madison, WI, pp. 49–58. https://doi.org/10.2134/1999.geneticsandexploitation.c5

Dias, L.A., Picoli, E.A., Rocha, R.B., Alfenas, A.C., 2004. A priori choice of hybrid parents in plants. Genet Mol Res 3, 356–368.

Dobzhansky, T., 1950. Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of Drosophila pseudoobscura. Genetics 35, 288–302.

Elson, H.A., 2011. Housing and husbandry of laying hens: past, present and future. Lohmann Inf. 46, 16–24.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., Jensen, P., Andersson, L., 2008. Identification of the Yellow Skin Gene Reveals a Hybrid Origin of the Domestic Chicken. PLOS Genet. 4, e1000010.

Fairfull, R.W., 1990. Heterosis. Poult. Breed. Genet. 913–933.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to Quantitative Genetics. Longman, Harlow.

Gavora, J.S., Fairfull, R.W., Benkel, B.F., Cantwell, W.J., Chambers, J.R., 1996. Prediction of heterosis from DNA fingerprints in chickens. Genetics 144, 777–784.

Goodnight, C.J., 1999. Epistasis and Heterosis, in: Coors, J.G., Pandey, S. (Eds.), The Genetics and Exploitation of Heterosis in Crops. American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., Madison, WI, pp. 59–68. https://doi.org/10.2134/1999.geneticsandexploitation.c6

Gosey, J., 2005. Crossbreeding the forgotten tool, in: Range Beef Cow Symposium. p. 32.

Haberfeld, A., Dunnington, E.A., Siegel, P.B., Hillel, J., 1996. Heterosis and DNA fingerprinting in chickens. Poult Sci 75, 951–953.

Hallauer, A.R., Miranda, J.B., 1988. Quantitative genetics in maize breeding. Iowa State University Press, Ames, IA.

Kress, D.D., Nelsen, T.C., 1988. Crossbreeding beef cattle for western range environments.

Krishnan, G.S., Singh, A.K., Waters, D.L.E., Henry, R.J., 2013. Molecular Markers for Harnessing Heterosis, in: Henry, R.J. (Ed.), Molecular Markers in Plants. GB: Wiley-Blackwell Ltd., Oxford, UK, pp. 119–136. https://doi.org/10.1002/9781118473023.ch8

Lamkey, K.R., Edwards, J.W., 1999. The quantitative genetics of heterosis, in: Coors, J.G., Pandey, S. (Eds.), The Genetics and Exploitation of Heterosis in Crops. American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., Madison, WI, pp. 31–48.

Lynch, M., Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates, Inc, Sunderland, MA.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157, 1819–1829.

Minvielle, F., Coville, J., Krupa, A., Monvoisin, J.L., Maeda, Y., Okamoto, S., 2000. Genetic similarity and relationships of DNA fingerprints with performance and with heterosis in Japanese quail lines from two origins and under reciprocal recurrent or within-line selection for early egg production. Genet Sel Evol 32, 289–302.

Muir, W.M., Cheng, H.-W., Croney, C., 2014. Methods to address poultry robustness and welfare issues through breeding and associated ethical considerations. Front. Genet. 5, 407. https://doi.org/10.3389/fgene.2014.00407

Neeteson-van Nieuwenhoven, A.-M., Knap, P., Avendaño, S., 2013. The role of sustainable commercial pig and poultry breeding for food security. Anim. Front. 3, 52–57. https://doi.org/10.2527/af.2013-0008

Nitter, G., 1978. Breed utilization for meat production in sheep, in: Animal Breeding Abstract. pp. 131–143.

Reif, J.C., Melchinger, A.E., Xia, X.C., Warburton, M.L., Hoisington, D.A., Vasal, S.K., Srinivasan, G., Bohn, M., Frisch, M., 2003. Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci. 43, 1275–1282.

Shull, G.H., 1952. Beginnings of the heterosis concept. Beginnings of the heterosis concept.

Shull, G.H., 1914. Duplicate genes for capsule-form in Bursa bursa-pastoris . Z Indukt Abstamm Vererbungsl 12, 97–149.

Thomson, A.L., 1964. A New Dictionary of Birds, Bird-Banding. New York McGraw-Hill. https://doi.org/10.2307/4511223

Van Sambeek, F., 2011. Breeding for 500 eggs in 100 weeks. World Poult. 27, 3.

Wolc, A., Kranis, A., Arango, J., Settar, P., Fulton, J.E., O'Sullivan, N.P., Avendano, A., Watson, K.A., Hickey, J.M., de los Campos, G., Fernando, R.L., Garrick, D.J., Dekkers, J.C.M., 2016. Implementation of genomic selection in the poultry industry. Anim. Front. 6, 23–31. https://doi.org/10.2527/af.2016-0004

Yamada, Y., 1988. The contribution of poultry science to society. Worlds. Poult. Sci. J. 44, 172–178. https://doi.org/DOI: 10.1079/WPS19880017

CHAPTER 6

General Discussion

6.1 Introduction

Crossbreeding is practiced extensively in commercial breeding programs of many plant and animal species, in order to exploit heterosis, breed complementarity, and to protect pure line genetic material. Because we still lack the knowledge to predict the performance of a cross, the decision on which combination of parental lines to use to make a cross is currently based on field testing of many potential crosses. However, as the number of pure lines increases, it becomes less feasible to test all possible crosses of the pure lines. The ability to accurately predict heterosis using information from the parental pure lines could therefore improve the efficiency of crossbreeding schemes by providing a basis on which to pre-select a subset of pure line combinations that can then be evaluated through field tests. Moreover, investigation of the genetic background of heterosis is also a relevant scientific question in its own right.

To this end, the research in this thesis focused mainly on the development of models to predict heterosis in White Leghorn crossbreds using genomic information from their parental pure lines. Based on a dominance model, we hypothesized that the genome-wide average of the squared difference in allele frequency (SDAF) at the SNP loci of the two parental lines might be a promising predictor of heterosis in the cross of these lines. Our results showed that the SDAF between parental pure lines is indeed a suitable predictor of heterosis in egg number and egg weight, with an accuracy of ~0.5 for our set of White Leghorn chicken lines (**Chapter 2**). We also showed that heterosis can be predicted at the individual sire level, using "heterozygosity excess" in the offspring of a sire, calculated from individual sire genotypes. In this way one can *in principle* further exploit the variation between sires from the same pure line, thereby maximizing the amount of heterosis expressed by the crossbreds. However, for the populations examined here, benefits were relatively limited (**Chapter 3**).

Because dominance effects may differ between loci, not all loci may contribute equally to heterosis. Therefore, in **Chapter 4**, we estimated variance components and additive and dominance effects of single nucleotide polymorphism (SNP) markers on egg number and egg weight in four White Leghorn pure lines, and discussed the possibility of using SDAF weighted by the estimated dominance effects of SNPs for the prediction of heterosis in their crosses. We found that dominance variance accounted for a relatively large proportion of the genetic variance in EN (~33%), but not in EW (~4%). In addition, the relative values of dominance effects

were much larger at some SNPs than at others, suggesting that some loci contribute much more to heterosis than others. Correlations between the raw SDAF and weighted SDAFs showed that prediction of heterosis based on a weighted SDAF would yield a considerably different ranking of crosses for each trait, compared with a prediction based on the raw SDAF. This implies that different lines would be selected for crossbreeding depending on the criterion used to predict heterosis.

In **Chapter 5**, we performed an exploratory genome-wide association study in order to gain insight into the genetic architecture of crossbred egg number and egg weight. We showed that egg number is a very polygenic trait controlled by at least ~1000 loci, and we identified several quantitative trait loci for egg weight.

In this **General Discussion,** I discuss the genomic prediction of heterosis, focusing on possible reasons for the lack of a consensus on the approach to predict heterosis, even after decades of research. I also suggest improvements for genomic prediction of heterosis, considering the advancements in genotyping and computation methods. Next, I give an example of the application of results from this thesis in crossbreeding programs.

6.2 Genomic prediction of heterosis

Several studies related to the prediction of heterosis have been done in the past on both plants and animals, however, there is no consensus on how to best predict heterosis (Atzmon et al., 2002; Balestre et al., 2009, 2008; Gavora et al., 1996; Haberfeld et al., 1996; Reif et al., 2003; Vuylsteke et al., 2000). In this section, I discuss possible reasons for the inability to reach a consensus on how best to predict heterosis, by reflecting on how heterosis was predicted in the past. I will address two main topics: 1) differences in methodology; 2) differences in the scientific merit of studies.

6.2.1 Differences in methodology

6.2.1.1 Predictor variables: squared difference in allele frequency (SDAF) versus genetic distance (GD)

Although the quantitative genetic theory linking heterosis to SDAF was published by Falconer as far back as 1960, prior to this thesis no studies directly testing this theory have been published. The theory shows that when heterosis is due to dominance,

the amount of heterosis due to a single bi-allelic locus is proportional to the SDAF between the two parental lines of the cross:

$$Heterosis_{ij} = (p_i - p_j)^2 d \quad \text{Eq. 6.1,}$$

where p_i and p_j are the allele frequencies at a particular locus in parental populations *i* and *j,* respectively, and *d* is the dominance deviation at that locus. A majority of the past studies on genomic prediction of heterosis mentioned this theory, but remarkably, none of them directly tested it. Instead, past studies used "genetic distance" (GD) as the predictor of heterosis. GD is a numeric measure of the extent of allele frequency difference or genetic divergence between species, populations or individuals, inferred from genetic markers (Nei, 1987, 1972). Examples of GD that are used frequently are Rogers' distance (Rogers, 1972), modified Rogers' distance (Wright, 1984), Cavalli-Sforza chord distance (Cavalli-Sforza and Edwards, 1967) and Nei's GD (Nei, 1972). Genetic markers commonly used in these studies on heterosis prediction are restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs) and microsatellites. These markers, which are multi-allelic, were used to compute GD and subsequently, the GD was used to predict heterosis.

How similar are genetic distances to SDAF? Do they have the same power to predict heterosis? To compare GD with SDAF, we computed pairwise correlations between SDAF and several measures of genetic distance based on 60K SNP allele frequencies, and found correlations between 0.98 – 1 (Chapter 2). We also compared the predictive ability of SDAF and the genetic distance with the lowest correlation to SDAF (Rogers' and modified Rogers' distance), and found almost identical results. This indicates that with a relatively large number of markers, SDAF and genetic distances calculated from *bi-allelic* markers have the same predictive ability for heterosis.

However, past studies used GD calculated from a limited number of *multi-allelic* genetic markers, and both the number and the type of marker may have had an effect on the similarity between GD and SDAF, and thus on predictive power. The effect of the number of markers on the prediction of heterosis is discussed in a later section.

In conclusion, if both the GD and SDAF are calculated from *bi-allelic* marker data, then the correlation between them is ~1, and thus I assume that they both have the same predictive power for heterosis. However, GD from *multi-allelic* markers may be

less correlated with SDAF, and therefore its power to predict heterosis could also be lower. It would be interesting to investigate this, because it might explain why past studies which used GD inferred from *multi-allelic* markers did not get high accuracies for the prediction of heterosis. One thing to mention however, is that these past studies may have opted for GD over SDAF because with *multi-allelic* markers, the definition of SDAF is not straightforward.

6.2.1.2 Target of predictions

In crossbreeding, the most important outcome is a crossbred production animal that meets the breeders' expectations – in other words, crossbred performance is what is important. For this reason, researchers would ultimately want to be able to predict crossbred performance.

There are two main models to partition crossbred performance. The first is a heterosis model:

$$\mu_{ij} = \frac{\mu_i + \mu_j}{2} + heterosis_{ij} \quad \text{Eq. 6.2,}$$

where μ_{ij} is the average phenotype of an *i*×*j* crossbred, $\mu_i \; and \; \mu_j$ are the average phenotypes of pure lines *i* and *j* respectively, and $heterosis_{ij}$ is the average heterosis expressed by an *i*×*j* crossbred. As can be seen from Eq. 6.2, heterosis is the deviation of the crossbred from the average of its two parental pure lines (Shull, 1952). Following from Eqs. 6.1 and 6.2, we have the following prediction for the mean phenotypic value of the crossbred:

$$\hat{\mu}_{ij} = \frac{\mu_i + \mu_j}{2} + \beta \cdot SDAF_{ij} \quad \text{Eq. 6.3.}$$

The second way to partition a crossbred phenotype is with a combining ability model:

$$\mu_{ij} = \mu_{SET} + GCA_i + GCA_j + SCA_{ij} \quad \text{Eq. 6.4,}$$

where μ_{ij} is the average phenotype of an *i*×*j* crossbred, μ_{SET} is an overall mean, the value of which depends on the set of crosses included in the analysis, $GCA_i \; and \; GCA_j$ are the general combining abilities of pure lines *i* and *j,* respectively, and SCA_{ij} is the specific combining ability of an *i*×*j* cross. The GCA is the average performance of a line in all its hybrid combinations (as a deviation from the overall mean, μ_{SET}), and SCA is the deviation of a particular hybrid combination from what would be expected on the basis of the average phenotype of all the hybrids

descending from its parental pure lines (Sprague and Tatum, 1942). Note that the GCA is not the same as the pure line mean.

One can see the similarity in the definitions of heterosis and SCA; however, their statistical and theoretical bases are very different. Statistically, GCAs are fitted as main effects, so that the average heterosis in all the hybrids descending from a pure line gets included in the GCA estimate of that line. The SCA is defined as a statistical interaction term, and the model constrains the SCA estimates to sum to zero. This automatically means that both GCAs and SCAs depend on the other crosses that are in the dataset.

On the other hand, heterosis does not depend on the other crosses in the dataset. In Chapter 2, we addressed this topic with a supplementary Excel sheet where we demonstrated that if heterosis is due to dominance, then an SDAF model (Eq. 6.3) partitions crossbred phenotypes into pure line averages and heterosis, whereas a GCA/SCA model does not. We also showed that predicted heterosis does not depend on which crosses are present in the dataset, whereas GCA and SCA estimates change depending on which other crosses are added/removed from the dataset being analysed. The dependency of GCA and SCA on the set of crosses included in the analysis hampers the comparison of experiments that partly include the same set of lines and/or crosses.

A heterosis model is therefore better suited to situations where new lines need to be evaluated continually. In addition, theory shows that heterosis is proportional to SDAF in the presence of directional dominance. SCA on the other hand is a complex function of additive and dominance effects and allele frequencies of the parental pure lines. This begs the question whether there is any theoretical justification for expecting genetic distance to be predictive of SCA, as several past studies have assumed? In Chapter 2, we showed that for egg number, the correlation between SDAF and SCA is considerably lower (0.3) than between SDAF and heterosis (0.6). This may be one of the reasons for the inconclusive results from past studies on the prediction of 'heterosis', because many of the studies were actually looking at SCA – not heterosis – and those two are not the same.

6.2.1.3 Measuring the accuracy of predicted heterosis

Another possible reason for the inconclusive results of studies on the prediction of heterosis is that different measures are used to assess the accuracy of predicted heterosis, and therefore one cannot clearly compare the outcomes of the various

studies in order to draw a conclusion. In my opinion, prediction accuracy should be assessed by the ability to predict crosses that were not part of the training dataset. For this reason, we performed a leave-one-out cross-validation in Chapter 2, where we removed all records of a particular cross from the data, and then predicted heterosis for the cross that had been left, out using the remaining data. We then took the correlation between observed and predicted heterosis as the measure of accuracy, and obtained a value of 0.6 for egg number, and 0.4 for egg weight. If we had instead taken the correlation between predicted heterosis based on the full data and observed heterosis as the measure of accuracy, we would have obtained an 'accuracy' of 0.7 for egg number and 0.6 for egg weight. Several of the past studies used correlations between the predictor based on the full data and observed heterosis or SCA as their measure of accuracy. This shows that the outcomes of different studies may not be directly comparable, making it difficult to draw conclusions based on reviewing past literature.

6.2.2 Differences in the scientific merit of studies

The scientific merit of a study depends on the type and amount of data, and how appropriate the methodology is for answering the scientific question at hand. For example, a study based on a large number of markers will probably give a more reliable estimate of SDAF or genetic distance than studies based on few markers. In this section, I will look at the effect of the number and informativeness of genetic markers on the accuracy of heterosis prediction.

6.2.2.1 Effect of the number of markers

In general, the accuracy of a marker-based predictor is affected by the level of linkage disequilibrium (LD) between the markers and underlying causative loci. For this reason, unless the causative loci themselves, or markers in high LD with them are known, one alternative would be to use a large number of markers spread densely across the entire genome, with the assumption that with such an extensive coverage of the genome, one would be able to capture the effect of the unknown underlying loci. Another perspective with more bearing on the prediction of heterosis is that with a larger number of markers, one gets a more accurate estimate of the true SDAF or genetic distance between parental pure lines, and that this genome-wide value also reflects the SDAF at the causative loci affecting the trait(s) of interest. These two lines of reasoning must be behind the conclusion by several authors (Dias et al., 2004; Krishnan et al., 2013; Rajendrakumar et al., 2015) that one

reason past studies on marker-based prediction of heterosis were inconclusive is that the number of markers used was too small.

Therefore, to test the effect of the number of markers on the accuracy of predicting heterosis, I investigated how the number of markers affects the estimate of the predictor variable, SDAF. For any two parental lines, say *i* and *j,* SDAF is calculated as follows:

$$SDAF_{ij} = \frac{\sum_{n=1}^{N}(p_{i_n} - p_{j_n})^2}{N} \quad \text{Eq. 6.5,}$$

where p_{i_n,j_n} is the allele frequency of SNP *n* in lines *i* and *j* respectively, and *N* is the total number of SNPs.

My "true" SDAF was the genome-wide average SDAF calculated from the full 60K SNP data, denoted as $SDAF_{60K}$. Since there were 45 different *i*×*j* combinations in my dataset, I had 45 $SDAF_{60K}$ values. Next, I created subsets of *N* = 200, 400, 800, 2000, 10K and 30K SNPs, selected randomly, but such that all chromosomes were equally represented, as far as possible (for example, chromosome 30 does not have many SNPs, so in some instances, even if all its SNPs were included, they were still fewer than the SNPs from chromosome 1). For each *N*, I repeated the SNP selection and estimation of $SDAF_N$ 100 times. For example, for the scenario with 200 SNPs, I obtained 100 different subsets each with 200 SNPs, and thus 100 estimates of $SDAF_{200}$ for each *i*×*j* combination.

Figure 6.1 shows a plot of the $SDAF_{60K}$ estimates against $SDAF_N$. It is clear that as the number of SNPs increases, the estimated SDAF gets closer to $SDAF_{60K}$. This shows that in general, as the number of SNPs increases, one is better able to estimate the true genome-wide level of divergence between populations. One can see that the estimates from 10K SNPs are almost as precise as those from 30K SNPs, which indicates that 10K genome-wide SNPs are probably sufficient to determine the divergence between the White Leghorn pure lines used in this analysis. Using less than 10K SNPs would result in a loss of accuracy. In addition, when the number of SNP dropped below ~1000, we found regression coefficients of observed ($SDAF_{60k}$) on predicted ($SDAF_N$) SDAF smaller than 1. This indicates a bias in predicted SDAF, where predictions overestimate the true differences between crosses in $SDAF_{60k}$.

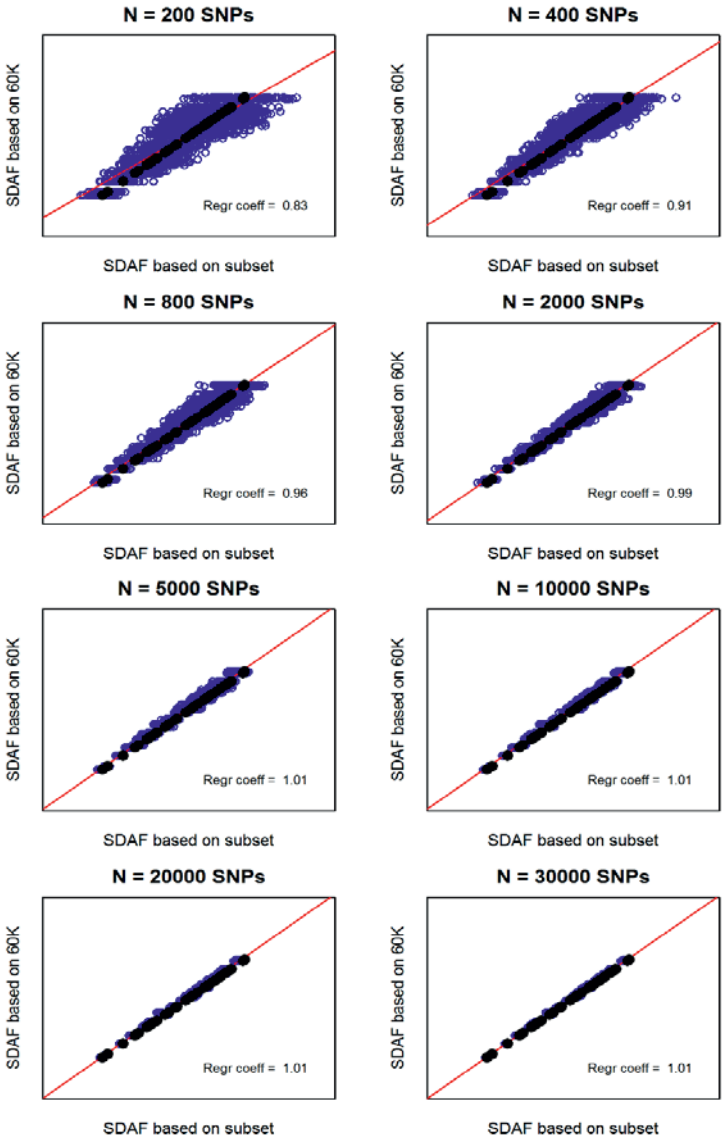

Figure 6.1 Plot showing estimates of the squared difference in allele frequency (SDAF) for 100 subsets and 45 different pure line combinations. In all graphs, the black points show SDAF based on 60K SNPs ($SDAF_{60K}$). The blue points show SDAF estimates from 100 subsets each of size N ($SDAF_N$). N is indicated in the titles of the sub-plots. The red line is the regression of $SDAF_{60K}$ on $SDAF_N$, and "Regr" coefficient is the resulting regression coefficient.

Figure 6.2 gives the standard deviation of SDAFs obtained from the 100 subsets for each N. This shows the amount of variation between the subsets; the larger the variation, the less reliable the estimated $SDAF_N$ is.

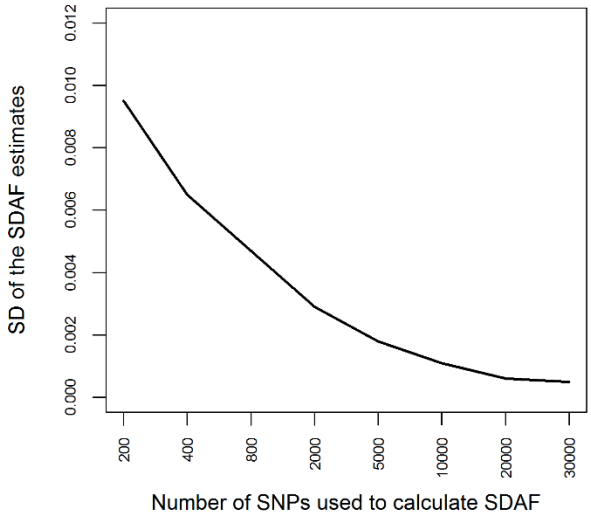

Figure 6.2 Plot of the standard deviation (SD) of the SDAF estimates obtained from using different numbers of SNPs.

The most important outcome of a heterosis prediction is the resulting rank of the crosses, because that is the basis of selection decisions. Therefore, to get a measure of how consistent the ranking of crosses was between the different subsets, I calculated Spearman's rank correlation coefficient between all the $SDAF_N$ and $SDAF_{60K}$. This would show whether crosses were consistently ranked in the same order irrespective of the number of SNPs used to calculate SDAF. Table 6.1 gives the results. Again, one can conclude that for this data, about 10K SNPs are enough to give the same ranking of crosses as the 60K SNPs.

Table 6.1 Spearman's rank correlation coefficients between $SDAF_{60K}$ and $SDAF_N$

Number of SNPs†	$\bar{r}_{SDAF_{60K},SDAF_N}$ (SD)
200	0.88 (0.05)
400	0.93 (0.03)
800	0.96 (0.02)
2000	0.98 (0.01)
5000	0.99 (0.004)
10 000	0.99 (0.002)
20 000	0.99 (0.001)
30 000	1.00 (<0.001)

†Number of SNPs in the subset used to estimate the squared difference in allele frequency (SDAF). $SDAF_N$ denotes and SDAF calculated from N number of SNPs
r = Spearman's rank correlation coefficient; SD = standard deviation

These results show that the number of markers indeed has a bearing on the estimation of SDAF (and/or genetic distances) and therefore, would affect the power to predict heterosis accurately. Deciding on the ideal number of SNPs to be used for future studies would depend upon the genome size – which is species-specific – as well as the diversity of the pure lines being evaluated. Based on the analyses above, I would recommend that future studies on laying hens should use at least 10K SNPs, or if using multi-allelic markers, then numbers that would give the same level of information as 10K SNPs should be used. For example, according to Schopen *et al.,* (2008), for each microsatellite marker, about three 3 SNPs are needed to obtain the same amount of information. This implies that about 3350 microsatellite markers would be needed for estimating SDAF in the example described here.

To my knowledge, the number of markers used in past studies on heterosis prediction was always below 700, which suggests that the estimated genetic distances were not sufficiently accurate for the prediction of heterosis.

6.2.2.2 Effect of the informativeness of markers

The accuracy of the prediction of heterosis may increase if a subset of markers that have been identified to have an effect on the trait of interest are used, instead of using all available markers. In principle, if all quantitative trait loci (QTL) affecting a

trait are known, then using information from a large number of markers that do not have an effect on the trait, or which are not in high LD with the QTL, may dilute the information from the QTL. On the other hand, if no prior information on QTL is known, perhaps using a relatively large number of SNPs could still be advantageous.

To investigate this issue, I extended the example given in section 6.2.2.1:
I randomly selected and omitted 2000 SNPs from the marker data and assumed that they were true QTL affecting the trait. I assumed that the SNPs on my chip are representative of the QTL. I then estimated SDAF based on only the QTL, $SDAF_{QTL}$, and calculated correlations between $SDAF_{QTL}$ and $SDAF_N$ from several subsets of different sizes (Table 6.2).

Results show that as the number of SNPs in the subset increased, the correlation between $SDAF_{QTL}$ and $SDAF_N$ also increased, implying that in situations where no prior information on QTL is known, using a relatively large number of SNPs to calculate SDAF is expected to give a more accurate estimate of the $SDAF_{QTL}$ than using a small number of SNPs. Take note however, that even though the correlation kept increasing as the number of SNPs increased, it never reached a value of 1. In addition, note that even with 30K SNPs, the correlation between $SDAF_{QTL}$ and $SDAF_N$ was only 0.98, whereas in the previous section (where no QTL were omitted from the data), I achieved a correlation of 0.98 with only 2K SNPs, and a correlation of 1 with 30K SNPs.

These results indicate that if QTL truly exist, then the advantage of adding extra SNPs which are *not* the QTL (or not in high LD with the QTL) is limited.

Table 6.2 Spearman's rank correlation coefficients between $SDAF_{QTL}$ and $SDAF_N$

Number of SNPs†	$\overline{r}_{SDAF_{QTL},SDAF_N}$ (SD)
200	0.88 (0.04)
400	0.92 (0.03)
800	0.95 (0.02)
2000	0.97 (0.009)
5000	0.98 (0.005)
10 000	0.98 (0.003)
20 000	0.98 (0.002)
30 000	0.98 (0.002)

†Number of SNPs in the subset used to estimate the squared difference in allele frequency (SDAF). $SDAF_N$ denotes an SDAF calculated from N number of SNPs. $SDAF_{QTL}$ is SDAF calculated from 2000 SNPs assumed to be true QTL.
r = Spearman's rank correlation coefficient; SD = standard deviation

Other authors have also written in support of using pre-selected subsets of SNPs for genomic predictions (Macciotta et al., 2009; Ober et al., 2015; Raymond et al., 2018), and more specifically for the genomic prediction of heterosis (Cho et al., 2004). However, research is still needed to determine the best criteria for selecting the appropriate subset of SNPs to be used. For example, whether to pre-select SNPs that have significant additive and/or dominance effects on the traits of interest – and if so, should these effects be estimated for single traits, composite traits or using a selection index?

Moreover, preselection of SNPs may be based on SNP effects that were estimated from either purebred or crossbred data. In general, one can say that if dominance variance is an important component of the phenotypic variance of the trait of interest, then is it is beneficial to use crossbred phenotypes in evaluations. Therefore, the decision on whether to use purebred or crossbred phenotypes (or both) for the estimation of SNP effects (which can then be used to weight SNPs for calculating SDAF) should not be taken lightly.

For heterosis due to directional dominance, it may be more important to identify SNPs that have positive estimated dominance effects, rather than additive effects. Even if so, one is still faced with the question of deciding how to use the dominance

effects that were estimated from the two pure lines that produced the cross. For example, in Chapter 4, for each locus, we used the average of the estimated dominance SNP effects from the two pure lines producing the cross to calculate the weighting factors for SDAF.

Therefore, as seen from this and the previous section, because in most situations all the true QTL are not known, one needs to reach a reasonable compromise between removing what are perceived to be 'uninformative' markers while still keeping a large enough number of markers to be representative of the genetic make-up of the individuals or population being evaluated.

6.2.3 Future prospects for the prediction of heterosis

With the current availability of dense genome-wide markers, and improvements in statistical modelling and computational ability, it is interesting to explore possibilities for improving the prediction of heterosis. According to theory, dominance is one of the main contributors to heterosis (Falconer and Mackay, 1996), therefore, once dominance effects can be estimated accurately, the next step is the development of heterosis prediction models that incorporate them appropriately.

Using SNP data and genomic selection methodology, it is now possible to create kinship matrices that can be used to disentangle additive and dominance effects, as well as epistatic effects (Vitezica et al., 2013). Dominance SNP effects can be estimated using a two-step approach. In the first step, genomic breeding values and animal dominance deviations are obtained from individuals that have been typed for SNPs and also recorded for the phenotype of interest. In the second step, the animal dominance deviations are back-solved to obtain estimated dominance effects of SNPs. We did this in **Chapter 4**, then used the estimated dominance effects to calculate weights for pairwise combinations of four White Leghorn pure lines. We found that there was a wide variation in the magnitude of weights assigned to the SNPs. These weights were further used to calculate a weighted genome-wide squared difference in allele frequency (WSDAF) between pure lines. Using WSDAF as a predictor would mean that certain SNPs contribute to the prediction of heterosis much more than others. Also, judging from the correlation between SDAF and WSDAF for egg number (-0.04) and egg weight (0.59) we concluded that predictions based on either SDAF or WSDAF would lead to very different selection decisions. We propose that a WSDAF model should be validated with real data

One benefit of being able to estimate dominance (and other non-additive effects) is that because the estimated effects will be trait-specific, the resulting heterosis predictions will also be trait-specific. This will be an improvement upon the current models that predict the same relative magnitude of heterosis irrespective of the trait (e.g Amuzu-Aweh et al., 2013), because phenotypic data clearly shows that heterosis is trait-specific: for example larger for egg number than for egg weight.

Another potential way to improve heterosis predictions is to find a way to differentiate between reciprocal crosses. Reciprocal crosses differ in their phenotypes (*e.g.* Peeters et al., 2012, this Thesis); however, SDAF (and the proposed dominance-weighted SDAF) has the limitation that it predicts the same expected heterosis for reciprocal crosses, *i.e* an A×B cross will get the same prediction as a B×A cross. In chickens, females are the heterogametic sex, therefore a female's Z chromosome is always inherited from its sire. The Z chromosome has been reported to have a parent-of-origin effect on survival (Peeters et al., 2012), and it may also have an effect on egg production traits. It would therefore be interesting to look into ways to incorporate information from the Z chromosome into heterosis predictions.

6.3 Including genomic prediction of heterosis in crossbreeding programs

New (pure) lines are introduced into breeding programs in several ways, for example breeders may develop new lines that are better adapted to new production conditions, or that meet new consumer demands. New lines will also be introduced after breeding companies merge, as has been the case in the history of Hendrix Genetics. Hendrix Genetics started off as a small farm in 1923, and over decades, several mergers and acquisitions of smaller breeding companies (see Figure 6.3) have led to the creation of a large company which currently controls about 40% of the global laying-hen breeding industry (excluding China).

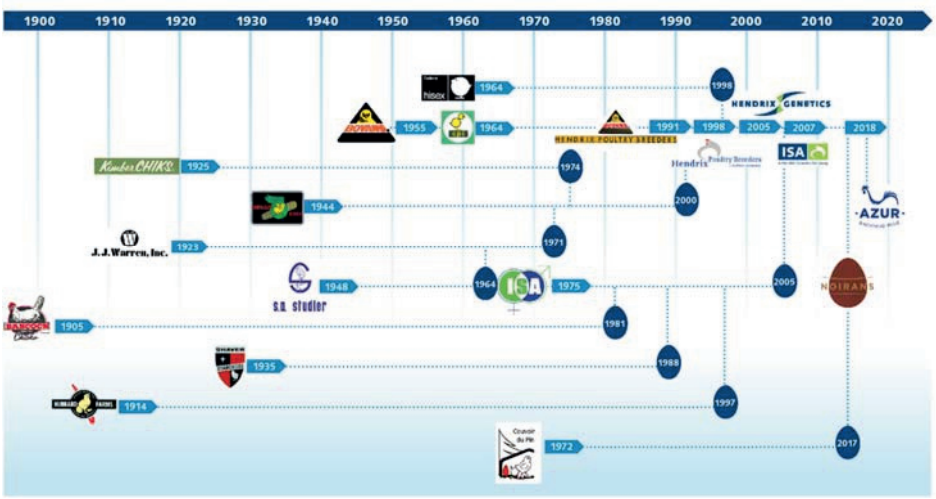

Figure 6.3 Mergers and acquisitions that led to the formation of the laying-hen division of Hendrix Genetics (used with permission of Hendrix Genetics).

Any time new lines are introduced into a breeding company, it is necessary to field-test them with the current lines and see if any desirable crossbred products could be made. If the possible crossbred products are many, then a pre-selection based on predicted heterosis could be used to reduce the number of crosses to be field-tested.

The fact that when using a heterosis model, new lines can be evaluated based solely on the genotypic information of the parental pure lines is a clear advantage over the general/specific combining ability model (G/SCA), because the G/SCA of a pure line can only be calculated *after* a field test has already been performed.

In **Chapter 2**, we showed that pre-selection based on predicted heterosis in egg number or egg weight could cut the number of crosses to be field-tested by up to 50%, with only ~ 4% loss in realised heterosis. These predictions were based on a raw genome-wide squared difference in allele frequency (SDAF), which had an accuracy of ~0.5. If the accuracy of prediction is increased, say, by improving the models with estimated non-additive effects, then the advantage could be even greater. In addition, the genomic prediction of heterosis could be relevant for plant breeding, where in principle, one can make an infinite number of pure lines by selfing – and thereby many potential hybrids could be made – way more than it is feasible to field-test. Predicted heterosis would therefore enable breeders to make an informed pre-selection of potential crosses to be field-tested.

Another instance where the genomic prediction of heterosis can be applied is at the onset of a breeding company or a national breeding scheme. Most developing countries have many diverse local breeds that are well-adapted to their environment and to the low-input extensive production system that is characteristic of the rural poultry sector. These local breeds are usually not well characterised, and neither is there any formal breeding scheme for them. There is a huge potential for improving the productivity of these local breeds, and judging from the advantages and success of crossbreeding in other parts of the world, perhaps developing countries could benefit greatly from starting an organized crossbreeding scheme. Crosses could be made between the local breeds or even by introducing high-producing foreign breed(s) in order to produce crossbreds that are still well-adapted to their environment, but have improved productivity.

A crossbreeding scheme however comes with increased complexity and may be more expensive than pure breeding, because all the breeds/lines involved in the crossbreeding scheme will each need to have their own breeding schemes. It is therefore important to perform a cost-benefit analysis to decide whether crossbreeding is the best option in the first place. In addition, the introduction of foreign breeds, if deemed necessary, must be done in an organized manner. If crossbreeding is decided upon, then obtaining SNP genotypes and calculating SDAFs between the selected breeds/lines could be one of the first steps in order to assess the genetic divergence among the breeds/lines and then pre-select potential crosses for field-testing.

6.4 Conclusions

The prediction of heterosis is a topic that has intrigued researchers for several decades. The findings herein have contributed to our knowledge on its prediction in White Leghorn crosses, and also added evidence that dominance is an important contributor to heterosis.

In addition, we estimated additive and dominance effects on egg number and egg weight in four White Leghorn pure lines, and proposed a method to incorporate the estimated dominance effects for the prediction of heterosis. We also reported genome-wide association results for crossbred egg number and egg weight, giving insight into the genetic architecture of these traits.

It would be interesting if the methods used in this thesis can be validated by studies in other populations of layers and other species where crossbreeding is practiced. I suggest that future studies should also focus on appropriate methods to include non-additive effects beyond dominance in the prediction of heterosis, and on how to predict reciprocal crosses.

6.5 References

Amuzu-Aweh, E.N., Bijma, P., Kinghorn, B.P., Vereijken, A., Visscher, J., van Arendonk, J.A., Bovenhuis, H., 2013. Prediction of heterosis using genome-wide SNP-marker data: application to egg production traits in white Leghorn crosses. Heredity (Edinb). 111, 530–8. https://doi.org/10.1038/hdy.2013.77

Atzmon, G., Cassuto, D., Lavi, U., Cahaner, U., Zeitlin, G., Hillel, J., 2002. DNA markers and crossbreeding scheme as means to select sires for heterosis in egg production of chickens. Anim.Genet. 33, 132–139.

Balestre, M., Machado, J.C., Lima, J.L., Souza, J.C., Nóbrega Filho, L., 2008. Genetic distance estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids. Genet Mol Res 7, 65–73.

Balestre, M., Von Pinho, R.G., Souza, J.C., Oliveira, R.I., 2009. Potential use of molecular markers for prediction of genotypic values in hybrid maize performance. Genet Mol Res 8, 1292–1306.

Cavalli-Sforza, L.L., Edwards, A.W.F., 1967. Phylogenetic Analysis: Models and Estimation Procedures. Evolution (N. Y). 21, 550–570. https://doi.org/10.2307/2406616

Cho, Y.-I., Park, C.-W., Kwon, S.-W., Chin, J.-H., Ji, H.-S., Park, K.-J., McCouch, S., Koh, H.-J., 2004. Key DNA Markers for Predicting Heterosis in F_1 Hybrids of *japonica* Rice. Breed. Sci. 54, 389–397. https://doi.org/10.1270/jsbbs.54.389

Dias, L.A., Picoli, E.A., Rocha, R.B., Alfenas, A.C., 2004. A priori choice of hybrid parents in plants. Genet Mol Res 3, 356–368.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to Quantitative Genetics. Longman, Harlow.

Gavora, J.S., Fairfull, R.W., Benkel, B.F., Cantwell, W.J., Chambers, J.R., 1996. Prediction of heterosis from DNA fingerprints in chickens. Genetics 144, 777–784.

Haberfeld, A., Dunnington, E.A., Siegel, P.B., Hillel, J., 1996. Heterosis and DNA fingerprinting in chickens. Poult Sci 75, 951–953.

Krishnan, G.S., Singh, A.K., Waters, D.L.E., Henry, R.J., 2013. Molecular Markers for Harnessing Heterosis, in: Henry, R.J. (Ed.), Molecular Markers in Plants. GB: Wiley-Blackwell Ltd., Oxford, UK, pp. 119–136. https://doi.org/10.1002/9781118473023.ch8

Macciotta, N.P.P., Gaspa, G., Steri, R., Pieramati, C., Carnier, P., Dimauro, C., 2009. Pre-selection of most significant SNPS for the estimation of genomic breeding values. BMC Proc. 3 Suppl 1, S14–S14. https://doi.org/10.1186/1753-6561-3-s1-s14

Nei, M., 1987. Molecular evolutionary genetics. Columbia university press.

Nei, M., 1972. Genetic Distance between Populations. Amer Nat 106, 283–292.

Ober, U., Huang, W., Magwire, M., Schlather, M., Simianer, H., Mackay, T.F.C., 2015. Accounting for genetic architecture improves sequence based genomic prediction for a Drosophila fitness trait. PLoS One 10, e0126880.

Peeters, K., Eppink, T.T., Ellen, E.D., Visscher, J., Bijma, P., 2012. Indirect genetic effects for survival in domestic chickens (Gallus gallus) are magnified in crossbred genotypes and show a parent-of-origin effect. Genetics 192, 705–713.

Rajendrakumar, P., K, H., Seetharama, N., 2015. Prediction of Heterosis in Crop Plants – Status and Prospects. Am. J. Exp. Agric. 9, 1–16. https://doi.org/10.9734/AJEA/2015/19263

Raymond, B., Bouwman, A.C., Wientjes, Y.C.J., Schrooten, C., Houwing-Duistermaat, J., Veerkamp, R.F., 2018. Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. Genet. Sel. Evol. 50, 49.

https://doi.org/10.1186/s12711-018-0419-5
Reif, J.C., Melchinger, A.E., Xia, X.C., Warburton, M.L., Hoisington, D.A., Vasal, S.K., Srinivasan, G., Bohn, M., Frisch, M., 2003. Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci. 43, 1275–1282.
Rogers, J.S., 1972. Measure of genetic similarity and genetic distance. Studies in genetics VII. Univ. Texas Publ. 7213, 145–153.
Schopen, G.C.B., Bovenhuis, H., Visker, M.H.P.W., Van Arendonk, J.A.M., 2008. Comparison of information content for microsatellites and SNPs in poultry and cattle. Anim. Genet. 39, 451–453. https://doi.org/10.1111/j.1365-2052.2008.01736.x
Shull, G.H., 1952. Beginnings of the heterosis concept. Beginnings of the heterosis concept.
Sprague, G.F., Tatum, L.A., 1942. General vs. specific combining ability in single crosses of corn 1. Agron. J. 34, 923–932.
Vitezica, Z.G., Varona, L., Legarra, A., 2013. On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195. https://doi.org/10.1534/genetics.113.155176
Vuylsteke, M., Kuiper, M., Stam, P., 2000. Chromosomal regions involved in hybrid performance and heterosis: their AFLP®-based indentification and practical use in prediction models. Heredity (Edinb). 85, 208–218.
Wright, S., 1984. Evolution and the genetics of populations, volume 4: variability within and among natural populations. University of Chicago press.

Summary

Summary

Heterosis is one of the most important benefits of crossbreeding. In situations where there are many different pure lines, breeders are faced with the challenge of identifying the best combinations of pure lines to produce crossbred products that express the best overall performance, which requires knowledge of heterosis. Currently, selection of parental pure lines is based on the results of field tests, during which the performance of their crossbred offspring is assessed under typical commercial settings.

Field tests are time-consuming, and also represent a large percent of the costs of commercial crossbreeding programs. This thesis therefore set out mainly to explore the possibilities and develop models for the accurate prediction of heterosis in White Leghorn crossbreds, using genomic information from their parental pure lines. Predicted heterosis could then be used to pre-select a subset of crosses to be assessed through field trials, thereby substantially reducing the costs of crossbreeding programs. We also hoped to gain insight into the genetic basis of heterosis. In addition, we explored the genetic architecture of egg number and egg weight in White Leghorns, both at the pure line and crossbred levels.

In **Chapter 2**, we studied egg number (EN), egg weight (EW) and survival days in 47 different White Leghorn crosses produced from 11 pure lines. Based on the theory that heterosis in a crossbred is proportional to the squared difference in allele frequency (SDAF) between its parental pure lines, we calculated a genome-wide squared difference in allele frequency (SDAF) between parental pure lines using 60K SNP genotypes. Results show that SDAF predicts heterosis in EN and EW at the line level with an accuracy of ~0.5, and that with this accuracy, one can reduce the number of field tests by 50%. We also showed that an SDAF model predicts heterosis whereas a combining ability model does not, which indicates that dominance is one of the important contributors to the genetic basis of heterosis. SDAF did not predict heterosis in survival days.

Moving beyond the line level, we aimed to predict heterosis at the individual sire level, in order to identify sires within the same (pure) line whose offspring would be superior in heterosis. Individual predictions would allow breeders to utilise the within-line genetic variation between sires, and potentially maximise heterosis in the offsping generation. Therefore, in **Chapter 3**, we derived the theoretical expectation of the amount of heterosis expressed by the offspring of an individual sire. Further, using 60K SNP genotypes of 3427 purebred sires and 16 types of crosses, we showed that individual sire genotypes can indeed be used to predict heterosis in their offspring. In our data however, the proportion of variation in genome-wide predicted heterosis due to sires from the same pure line was small (0.7%); most differences were observed between lines (99.0%). This led us to conclude that considering the genotyping costs involved, prediction of heterosis for individual sires would only be beneficial if sire genotypes are already available.

Quantitative genetic theory shows a clear proportionality between the dominance effect at a locus, SDAF and heterosis. This theory made us curious to explore the possibility of using dominance effects to improve the prediction of heterosis. Thus, in **Chapter 4**, we used 60K SNP genotypes and phenotypes of 11,119 females from four White Leghorn pure lines to estimate variance components, breeding values and dominance deviations for EN and EW. We then back-solved the dominance deviations to obtain estimated dominance effects of the SNPs. Next, we calculated a dominance-weighted SDAF for each trait. Our expectation was that a dominance-weighted SDAF will give trait-specific – and possibly more accurate – heterosis predictions than a raw genome-wide average SDAF.

We found that dominance variance accounted for up to 37% of the genetic variance in EN, and up to 4% of that in EW. Results showed that for both EN and EW, negative and positive estimated dominance effects are spread rather evenly across the genome. The relative values of the dominance effects were much larger at some

SNPs than at others, suggesting that some loci contribute much more to heterosis than others. We also found that the weighted SDAF for EN and EW were substantially different and showed greater variation than the raw SDAF, suggesting that a dominance-weighed SDAF may indeed have the potential to predict trait-specific heterosis. In addition, the correlations between the raw SDAF and the weighted SDAFs showed that prediction of heterosis based on a weighted SDAF would yield considerably different ranking of crosses for each trait, compared with a prediction based on the raw SDAF. This implies that different crosses would be selected depending on the criterion used to predict heterosis. These results justify further investigation into the application of a dominance-weighted predictor of heterosis.

In order to gain insight on the genetic architecture of crossbred EN and EW, in **Chapter 5**, we performed genome-wide association studies on EN and EW in a total of 16 commercial crossbreds, first using data from all crosses, and then for selected subsets. We found that EN is a highly polygenic trait controlled by at least a thousand loci, and that no large quantitative trait loci are segregating in the commercial White Leghorn crosses that we studied. For EW, we found that a few relatively large QTL are segregating in the population. This may be because EN has been under intense directional selection for several decades, whereas EW has been under less-intense, stabilising selection.

Finally, in the general discussion of this thesis (**Chapter 6)**, I discuss the genomic prediction of heterosis, focusing on possible reasons for the lack of a consensus on the approach to predict heterosis, even after decades of research. I also discuss new opportunities for the genomic prediction of heterosis, considering the advancements in genotyping and computation methods. Lastly, I give an example of the application of results from this thesis in crossbreeding programs.

The findings in this thesis have contributed to our knowledge on the prediction of heterosis in White Leghorn crosses, and also added evidence that dominance is an important contributor to heterosis. In addition, our results give insight into the genetic architecture of egg number and egg weight in several pure line and crossbred populations.

Sammanfattning

(Swedish summary)

Korsningseffekten, som även kallas för heterosis, är en av dem viktigaste effekterna av korsavel. Heterosis uppnås genom att korsa två rena raser och innebär att avkomman i genomsnitt har bättre egenskaper än föräldrarna. Metoden används för avel av flera olika djurslag, bland annat värphöns, som i den här avhandlingen.
När uppfödare har tillgång till flera renrasiga linjer är det en utmaning att identifiera den bästa kombination av raser som leder till en korsningseffekt som i sin tur resulterar i optimala egenskaper. Den här processen kräver kunskap om heterosis. För närvarande baseras urvalet av raserna för korsavel på fältexperiment där man bedömer prestationen av korsningarna under typiska kommersiella förutsättningar.

Fältförsök är tidskrävande och innebär även en stor kostnad för kommersiella program inom korsavel. Det övergripande syftet av den här avhandlingen är därför att både undersöka korsningsavelns möjligheter samt att utveckla modeller för att kunna förutsäga korsningseffekten hos kycklingrasen Vit Leghorn med hjälp av genomisk information från den renrasiga föräldragenerationen. Den förutsagda korsningseffekten kan sedan användas för att göra ett första urval bland möjliga korsningar som kommer att bedömas i fältförsök. Därmed skulle man kunna reducera kostnaden av korsavelsprogram. Vi hoppas även att få mer insikt i de genetiska förutsättningarna av korsningseffekten. Dessutom har vi undersökt den genetiska arkitekturen bakom antalet och vikten av ägg hos Vit Leghorn, både vad det gäller renrasiga och korsade linjer.

I **kapitel 2** har vi undersökt antalet ägg (egg number/EN), äggens vikt (egg weight/EW) och antal överlevnadsdagar i 47 olika korsningar från 11 renrasiga linjer av Vit Leghorn. Vi utgår ifrån teorin att mängden av heterosis i en korsning är proportionellt till den kvadratiska skillnaden i allel frekvenser mellan föräldrar linjer (s.k. SDAF). Vi skattade SDAF mellan alla 11 renrasiga linjer på hela genomet med hjälp av 60,000 SNP genotyper. SNP står för single nucleotide polymorphism – ”enbas-polymorfi” och används som en genetisk markör för variation mellan

individer. Resultaten visar att värdet för SDAF förutsäger korsningseffekten för antalet ägg (EN) och äggens vikt (EW) med en statistisk säkerhet av ~0.5. Med hjälp av dessa resultat kan fältförsöken sedan halveras. Vi visar också att en modell som använder SDAF-värdet kan förutsäga korsningseffekten medan en alternativ korsning modell som kallas för ”combining ability” (kombinations potential) inte kan göra detta. Detta tyder på att dominans är en viktig faktor för genetiken bakom korsningseffekten. SDAF kunde inte förutsäga någon korsningseffekt på antal överlevnadsdagar.

För att kunna förutsäga korsningseffekten i mer detalj ville vi i nästa steg identifiera renrasiga fäder som skulle ge upphov till en utmärkt korsningseffekt hos avkomman. Individuella förutsägelser skulle kunna göra det möjligt för uppfödare att använda den genetiska variationen som finns bland fäder inom samma ras, och därmed maximera korsningseffekten i nästa generation. Därför härleder vi i **kapitel 3** den teoretiska förväntade korsningseffekten i avkomman av en individuell fader. Genom att använda 60K SNP genotyper av 3427 renrasiga fäder och 16 typer av korsningar visar vi att genotypen av individuella fäder kan användas för att förutsäga korsningseffekten i avkomman. Andelen av variation i förutsägelsen av korsningseffekten som beror på fäder från samma linje är dock liten (0,7%); de flesta skillnader kunde observeras mellan olika linjer (99,0%). Med tanke på kostnaden för individuell genotypning är vår slutsats därför att förutsägelse av korsningseffekten på grund av individuella fäder är enbart av fördel om genotypen av fadern är redan tillgänglig.

Kvantitativ genetisk teori visar en tydlig proportionalitet mellan dominanseffekten vid ett genetisk lokus, SDAF och korsningseffekten. Vi ville gärna utforska möjligheten att använda dominanseffekter för att förbättra förutsägelsen av korsningseffekten. I **kapitel 4** har vi använt 60K SNP genotyper och fenotyper från 11119 honor ifrån Vit Leghorn renrasiga linjer för att uppskatta varianskomponenter

, avelsvärden och avvikelse pga dominans (dominance deviations) för antalet och vikten av äggen (EN och EW). Vi härledde sedan avvikelse pga dominans för att få uppskattningar av dominanseffekten av SNPar. Därefter räknade vi ut dominans-viktade SDAF för varje egenskap. Vi förväntade oss att en dominans-viktad SDAF borde ge en mer egenskapsspecifik - och därmed mere exakt - förutsägelse för korsningseffekten än ett genomsnittlig SDAF som baseras på hela genomet.

Vi upptäckte att varians pga dominans är ansvarig för upp till 37% för den genetiska variationen i antal ägg (EN) och 4% för den genetiska variation bakom äggens vikt (EW). Resultaten visar att negativa och positiva dominanseffekter är fördelade jämt över genomet, både vad det gäller äggens antal (EN) och vikt (EW). De relativa värden av dominanseffekten var mycket större vid vissa SNPar än andra, vilket tyder på att vissa loci (områden i arvsmassan) bidrar mer till korsningseffekten än andra. Vi upptäckte också att att de viktade SDAF för antalet och vikten av äggen (EN och EW) var väsentligt olika och visade en större variation än den vanliga SDAF, vilket tyder på att en dominans-viktad SDAF kan faktiskt ha potential att förutse egenskapsspecifika korsningseffekter. Dessutom visar korrelationerna mellan vanliga och viktade SDAF att förutsägelser baserade på den viktade SDAF skulle kunna ge en betydlig annorlunda rankning av korsningar för varje egenskap, jämfört med en förutsägelse som baseras på vanlig SDAF. Detta betyder att de olika korsningar skulle selekteras beroende på kriteriet som används för att förutse korsningseffekten. Resultaten rättfärdigar ytterligare undersökning av tillämpningen av dominans-viktad förutsägelse av korsningseffekten.

För att få insikt i den genetiska arkitekturen bakom EN och EW i korsavlade värphöns, genomförde vi i **kapitel 5** helgenom-associations studier på EN och EW i totalt 16 kommersiellt korsavlade raser. Vi använde först data från alla korsningar och därefter utvalda delar. Vi upptäckte att EN är till en hög grad en polygenetisk egenskap (en egenskap som beror på flera genetiska faktorer) som kontrolleras av

minst tusen gener, och att inga så kallade stora QTL (quantitative trait loci – regioner av DNA som har siknifikant effekt på kvantitative egenskaper) segregerar i korsningarna av Vit Leghorn som vi har studerat. För EW upptäckte vi att relativt få stora QTL segregerar i populationen. Detta kan bero på den intensiva selektionen för äggens antal (EN) under flera decennier, medan mindre selektion har gjorts för EW.

Kapitel 6 innehåller den övergripande diskussionen av den här avhandlingen och jag diskuterar den genetiska förutsägelsen av korsningseffekten, med fokus på möjliga anledningar för bristen av konsensus på tillvägagångssätt för att förutse korsningseffekten även efter flera decennier av forskning. Jag tar också upp nya möjligheter för genetisk förutsägelse av korsningseffekten, särskild med tanke på framstegen inom genotypning och beräkningsmetoder. Till sist ger jag ett exempel av tillämpningen av resultaten i den här avhandlingen inom korsavel.

De vetenskapliga fynden i den här avhandlingen har bidragit till kunskap om förutsägelsen av korsningseffekten i korsningar av kycklingrasen Vit Leghorn, och har bidragit med yterligare evidens att dominans är en viktig faktor för korsningseffekten. Dessutom ger våra resultat insikt i den genetiska arkitekturen bakom äggens antal och vikt i flera renrasiga linjer och korsade populationer.

CURRICULUM VITAE

About the author

Publications

Education and training

About the author

Esinam Nancy Amuzu-Aweh was born on the 2nd of January 1986, in Accra, Ghana. She obtained her basic education at Sol Plaatje Primary school in South Africa, and junior high education at Englebert School in Accra, Ghana. In 2000, Nancy started senior high school at St. Mary's Senior High in Accra, where she studied General Science. After her first year, she was adjudged the best science student, and received a scholarship for the rest of her high school education. In 2004, Nancy gained admission to the University of Ghana, Legon, and in 2008, she graduated with a BSc. (Hons) in Zoology. Nancy's thesis was on the phenotypic characterisation of cowpea weevils. After graduation, Nancy worked as a research assistant at the University of Ghana, on a crossbreeding project for local cultivars of cowpea. In August 2009, she won a scholarship to pursue the Erasmus Mundus Master's degree in Animal Breeding and Genetics. Nancy spent the first year of her Master's at the University of Natural Resources and Applied Life Sciences, Austria, and her second year at the Norwegian University of Life Sciences. She obtained her MSc degree in June 2011, with a thesis entitled 'Comparison of methods for estimating the effects of casein SNPs on milk traits in Norwegian goats'. In September 2011, Nancy was awarded a scholarship to pursue a joint PhD degree at Wageningen University, the Netherlands, and the Swedish University of Agricultural Sciences, under the European Graduate School in Animal Breeding and Genetics. The results of her PhD research are presented in this thesis. From Feb 2016 to September 2018, Nancy worked as a research fellow on a USAID project with University of California, Davis, Iowa State University, Sokoine University of Agriculture, Tanzania, and the University of Ghana. The project aimed at increasing productivity of local chicken breeds, and improving resistance to Newcastle disease. Nancy's main research interest is the quantitative genetics and genomics of crossbreeding.

Peer-reviewed publications

Amuzu-Aweh, E.N., Bovenhuis, H., de Koning, D.-J., Bijma, P., 2015. Predicting heterosis for egg production traits in crossbred offspring of individual White Leghorn sires using genome-wide SNP data. Genet. Sel. Evol. 47, 27. https://doi.org/10.1186/s12711-015-0088-6

Amuzu-Aweh, E.N., Bijma, P., Kinghorn, B.P., Vereijken, A., Visscher, J., van Arendonk, J.A., Bovenhuis, H., 2013. Prediction of heterosis using genome-wide SNP-marker data: application to egg production traits in White Leghorn crosses. Heredity (Edinb). 111, 530–8. https://doi.org/10.1038/hdy.2013.77

Walugembe, M., Mushi, J.R., **Amuzu-Aweh, E.N.,** Chiwanga, G.H., Msoffe, P.L., Wang, Y., Saelao, P., Kelly, T., Gallardo, R.A., Zhou, H., others, 2019. Genetic Analyses of Tanzanian Local Chicken Ecotypes Challenged with Newcastle Disease Virus. Genes (Basel). 10, 546.

Asante, I.K., **Amuzu, E.N.,** Donkor, A., Annan, K., 2009. Screening of some Ghanaian medicinal plants for phenolic compounds and radical scavenging activities. Pharmacogn. J. 1, 201–206.

Manuscripts in preparation

Amuzu-Aweh, E.N., Bijma, P., Calus, M. P. L., Bovenhuis, H.
Genomic estimation of variance components and dominance SNP effects for egg number and egg weight in White Leghorn pure lines

Amuzu-Aweh, E.N., Bijma, P., Bovenhuis, H.
A genome-wide association study for egg number and egg weight in a large crossbred population of White Leghorns

Conference proceedings

Amuzu-Aweh, E.N., Bijma, P., Kinghorn, B.P., Vereijken, A., Visscher, J., van Arendonk, J.A., Bovenhuis, H.,2013. Genomic prediction of heterosis for egg production traits in White Leghorn crosses. 64th Annual Meeting of the EAAP, Nantes, France

Amuzu-Aweh, E.N., Bovenhuis, H., de Koning, D.-J., Bijma, P., 2014. Prediction of Heterosis in White Leghorn Crossbreds using Paternal 60K SNP Genotypes. 10th World Congress of Genetics Applied to Livestock Production (WCGALP), Vancouver, Canada

Amuzu-Aweh, E.N., Bovenhuis, H., de Koning, D.-J., Bijma, P., 2014. Sire-based genomic prediction of heterosis in White Leghorn crossbreds. 65th Annual Meeting of the EAAP, Copenhagen, Denmark

Amuzu-Aweh, E.N., Bijma, P., Bovenhuis, H., de Koning, D.-J., 2015. Using loci with identified dominance effects to improve the prediction of heterosis. 66th Annual Meeting of the EAAP, Warsaw, Poland.

Amuzu-Aweh, E.N., Kayang, B., Muhairwa, A., Botchway, P., Naazie, A., Anning, G., Gallardo, R., Kelly, T., Zhou, H., Lamont, S., Dekkers, J., 2018. Genetic parameters and genomic regions associated with growth rate and immune response to Newcastle disease in local chicken ecotypes in Ghana and Tanzania. 11th World Congress of Genetics Applied to Livestock Production (WCGALP), Auckland, New Zealand

Dekkers, J., Botchway, P.K., **Amuzu-Aweh, E.N.**, Naazie, A., Aning, G., Zhou, H., Dekkers, J., Lamont, S., Gallardo, R., Kelly, T., Bunn, D., Kayang, B., 2018. Genotypic and phenotypic characterisation of three local chicken ecotypes of Ghana based on principal component analysis and body measurements. 11th World Congress of Genetics Applied to Livestock Production (WCGALP), Auckland, New Zealand

Kayang, B., **Amuzu-Aweh, E.N.,** Botchway, P., Tudeka, C., Naazie, A., Aning, G., Dekkers, J., Lamont, S., Gallardo, R., Kelly, T., Bunn, D., Zhou, H., 2018. Performance of three local chicken ecotypes of Ghana naturally exposed to velogenic Newcastle disease virus. 11th World Congress of Genetics Applied to Livestock Production (WCGALP), Auckland, New Zealand

Walugembe, M**., Amuzu-Aweh, E.N.,** Kayang B.B., Muhairwa, A.P., Botchway, P.K., Mushi, J. R., Honorati, G., Naazie, A., Aning, G., Msoffe, P., Wang, Y., Saelao, P., Kelly, T.R., Gallardo, R., Zhou, H., Lamont, S. J., Dekkers, J. C. M.,2019. Genetic Analyses of Ghana and Tanzania Local Chicken Ecotypes Challenged with Newcastle Disease Virus. Plant and Animal Genome XXVII Conference, San Diego, California

Ahiagbe, K. M. J., **Amuzu-Aweh, E. N.,** Avornyo, F.K., Adenyo, C., Nyame-Asem, J. K., Bonney, P., Amoah, K. O., Naazie, A., Kayang, B.B.,2019. Comparison of early growth and survivability in indigenous guinea fowls from Northern Ghana. 7th All Africa Conference on Animal Agriculture, Accra, Ghana

Tudeka, C.K., Kayang, B. B., Aning, K. G. , Naazie, A., Botchway P. K**., Amuzu-Aweh E. N.,** Enyetornye, B., Sarkwa, F. O., Kelly, T. R., Gallardo R., Zhou H., Fiadzomor, D., 2019. Response of three Ghanaian local chicken ecotypes to emsogenic and velogenic newcastle disease virus challenge. 7th All Africa Conference on Animal Agriculture, Accra, Ghana

EDUCATION AND TRAINING		
The Basic Package (9 ECTS)	**Year**	**Credits***
WIAS Introduction Course	2011	1.5
Ethics and philosophy of life sciences	2011	1.5
Welcome to EGS ABG (Paris)	2011	2.0
Summer Research School "Animal breeding and society"	2012, 2013	4.0
Scientific Exposure	**Year**	**Credits**
International conferences (5.4 ECTS)		
64th EAAP Annual Meeting, Nantes (France)	2013	1.2
65th EAAP Annual Meeting, Copenhagen (Denmark)	2014	1.2
10th World Congress on Genetics applied to Livestock Production(WCGALP), Vancouver (Canada)	2014	1.5
11th World Congress on Genetics applied to Livestock Production (WCGALP), Auckland (New Zealand)	2018	1.5
Seminars and workshops (1.2 ECTS)		
WIAS Science Day, Wageningen	2012,13	0.6
Hendrix Genetics Academy	2012	0.3
Genetics of Social Life: Agriculture meets Evolutionary Biology	2013	0.3
Presentations (5 ECTS)		
Hendrix Genetics Academy(Netherlands) ORAL	2012	-
64th EAAP Annual Meeting, Nantes (France) ORAL	2013	1.0
65th EAAP Annual Meeting, Copenhagen (Denkmark) ORAL	2014	1.0
10th World Congress on Genetics applied to Livestock Production (WCGALP), Vancouver (Canada) ORAL	2014	1.0
66th EAAP Annual Meeting, Warsaw (Poland) ORAL	2015	1.0
11th World Congress on Genetics applied to Livestock Production (WCGALP), Auckland (New Zealand) ORAL	2018	1.0
In-Depth Studies	**Year**	**Credits**
Disciplinary and interdisciplinary courses (14 ECTS)		
Advanced methods and algorithms in animal breeding with focus on genomic selection	2012	1.5
Identity by Decent (IBD) Approaches to Genomic Analyses of Genetic Traits	2012	1.2
Sequence Data Analysis Training School	2012	1.5
Social Genetic Effects: Theory and Genetic Analysis	2013	0.9
Advanced quantitative genetics for animal breeding	2014	3.0
Genetic Analysis unsing ASReml 4.0	2014	1.5
Introduction to theory and implementation of genomic selection	2014	1.35
Genomic Selection in Livestock	2015	1.5
Design of Breeding Programs with Genomic Selection	2015	1.5
Advanced statistics courses (6 ECTS)		
Statistics for the Life Sciences	2012	2.0
Introduction to Statistical Methods in Quantitative Genetics and Breeding	2014	4.0
PhD students' discussion groups (1.5 ECTS)		
Quantitative Genetics Discussion Group	2011 - 2015	1.5
Professional Skills Support Courses (5.5 ECTS)	**Year**	**Credits**
WGS Course: Techniques for Writing and presenting a scientific paper	2012	1.2
WGS Course: Writing grant proposals	2015	2.0
High-impact writing in science	2015	1.3
WGS Course: Teaching and Supervsing Thesis students	2012	1.0
Research Skills Training (2.6 ECTS)	**Year**	**Credits**
External training period Swedish University of Agricultural Sciences, Sweden	2014	2.0
Getting Started in ASReml	2012	-
Introduction to R for Statistical Analysis	2012	0.6
Didactic Skills Training	**Year**	**Credits**
Supervising practicals and excursions (1 ECTS)		
Animal Breeding and Genetics Course - WUR	2012	1.0
Education and Training Total		**51.2**

* one ECTS credit equals a study load of approximately 28 hours

Acknowledgements

My PhD journey has been great, but very different from the average. I am extremely grateful for the wonderful support system that has been by my side each step of the way.

Thank you Lord for life, health, the opportunity to pursue this PhD degree, and the ability to complete it.

Thanks to the members of my supervisory team for providing me with the knowledge, skills, encouragement and guidance needed to finalise this thesis.

Piter, in the beginning, I was scared that you were too smart for me, and that it would be difficult to keep up with your quick and analytical mind. I still think you are too smart ☺, but what I quickly realized is that you have the ability to immediately identify the strengths and weaknesses of your students and in that way you were able impart knowledge to me, while stimulating me to think independently and challenge myself each and every day. I am very grateful for all your time, support, patience, ideas, encouragement, kindness and even the blunt criticism when needed.

Henk, I cannot begin to list all the things I learnt from you over these years, but two big things I will take away are how important it is to fully understand and diagnose a model before looking at the main results it gives, and the importance of paying attention to the "tiny" details. Thank you for all your time, support, advice, encouragement, kindness and for teaching me so much about modelling.

Piter and Henk, I am especially grateful for the fact that you made me a priority during my last visit to Wageningen in 2019, and afterwards when I returned to Ghana to complete my thesis. You two surely make a great supervisory duo!!

Acknowlegments

DJ, thank you for all your support during my year in Sweden. It was a tough year for me, because I could not complete what I had planned to that year, but I later realised that it was during that year that I learnt so much! Thank you for all the discussions we had, your support, encouragement, and for teaching me the DJ-way of peeling an orange! Thanks also for the extra effort you put in to make sure all my paperwork was complete at the very end of my journey.

Piter, Henk and DJ, you three always welcomed me with a smile even when I had to come and see you outside our planned times. I know it has not been easy supervising me since I left Europe – all the e-mails and Skype calls – I really appreciate the effort it has taken to be there for me. Thank you!

Johan, thank you very much for the support you gave me throughout my journey – both scientific and administrative. I am grateful that I had such an open-minded and understanding promotor, and it is bitter-sweet that you are no longer my promotor, but Henk is great too, so I am blessed to have had two promotors during my PhD period.

To the entire R&D team at Hendrix Genetics, I am very grateful for your interest in my project, for receiving me warmly any time I visited, and for your input and suggestions during our meetings. Special thanks to Jeroen, Addie, Aline and Katrijn. I appreciate the time you took to prepare and send me data, respond to my queries, and read my manuscripts. Teun, thanks for showing me the ropes during my visit to Canada; it was great to experience things on the farm. Gerard and Marco, thank you for your interest in my project.

To all the staff, researchers and PhD candidates at ABGC, I could not have asked for a more stimulating environment. I learnt a lot during all the discussions at QDG, TLM,

coffee breaks, or when I just popped into your offices with questions. Special thanks to Mario, Alex, and Jeremie for help with programming-related issues.
You have all helped shape me into the budding researcher that I am today, and I look forward to collaborations in the future. I also thank members of ABGC Partycom for organising all the recreational activities we did together to de-stress and get to know each other better.

Lisette, you are always ready to offer assistance with a smile. Thank you for taking care of all the administrative things for me, and for being a friendly ear. You always go that extra mile to make sure things run smoothly. Thank you very much.

Thanks to Ada, Monique and Fadma from ABGC, and Helena, Harriet, Monica and Fernando from SLU for helping me with administrative paperwork.

Thanks to the coordinators, administrators, secretaries and all members of the "EGS-ABG" consortium for working so hard to make the program a success.
Thanks to the European Union and Hendrix Genetics for funding my PhD studies.

To all my friends, you surely made this PhD journey so much fun!! Marzi, you have been a very dear friend to me; we talk about anything! Thank you so much for being like a sister to me, and for all the encouraging words when times were a bit rough during my PhD journey. I know our friendship will continue to blossom for life. André and Sandrine, we grew from office-mates to become such good friends. I appreciate all the brainstorming sessions, chats, fun days, singing (with passion!) … so many lovely memories. André, what can I say? You were (still are) my bro! Work, squash courts, recording great music, all the good food, just "scratching", all the fun times really made my journey that much better. Also, just as we told ourselves many years ago, you are going to be my paranymph! Thank you. Marcos, you know you only ended up being my second-favourite Brazillian guy because I met André before you!

Acknowlegments

Thanks for all the discussions, dinners and fun times we had together. Big G, you are the only reason I managed to forgive Uruguay for what happened to Ghana in 2010. You are such a great person. Thanks for being my friend.
To the wonderful group pf PhD candidates that I spent time with at ABGC, thanks for being a friendly, lively group! We had lots of fun together. Ewa, Katrijn, Mahlet, Hamed, Hadi, Setegn, Juanma, Gus, Yogesh, Jovana, Mathieu, Claudia, Yvonne, Maria, Hooiling, Panya, Troncg, Susan, Dianne, Mirte, Tessa, Sonia, Anoop, Maulik, Vinicius, Coralia, Amabel, Mathijs, Mandy, Wosseni, Robert, Kip, Anouk, Aniek and Naomi. It was nice knowing you all.
Josie, Rashida and Nina, thanks for your friendship, support and encouragement during our time together in Wageningen! Josie, you were there for me at a very important time, thank you.

Many thanks to the SLU "PhD mafia"! Despite the weather, Sweden was always warm and bright because of you. Agnese, Chrissy, Ahmed, Bingjie, Merina, Thu, Shizhi, Xiaowei, Berihu, Alberto. We had so much fun together, in and out of work.

Thank you to all the members of HGEN at SLU, I enjoyed all our group discussions and "fiika" we had together.

Chrissy, Jovana, Biaty, Amaami, Paa Kwasi and Misghina, thanks for making my last visit to Wageningen a pleasurable one. Chrissy, thanks for being my paranymph.
Thank you Alex, Rosilde and Ina for happily sharing your office with me.

Auntie Monica and Uncle Steve, thanks for being my home in the Netherlands, God bless you! Thank you to WUR Sikaman, it was great having so many brothers and sisters in Wageningen.

To my entire family, I dedicate this thesis to you, because I surely could not have done it without you. Mama and daddy, I love you. Thank you for raising me to believe in myself, and thanks for all the love, support and encouragement. Mama, I know your prayers covered me all those years in Europe. Thank you. Abla and Dela, you are the best sisters ever! Thanks for always being there for me, and for setting good examples for me. Doree, thanks for always having my back. Suwa, bro-in-law "baako p3", thank you. To Wonder, Shed, Eunice, Lookie and Edem, thanks for always pushing me with "so when can we start calling you doc?". Francisca, thanks for everything.

Samuel, my love, you are my support, confidant, best friend, everything! Thank you thank you thank you!! Thanks also for being such a wonderful father to our boys; it made it possible for me to breathe easy during the periods I had to leave them behind. I love you!

To my two boisterous boys, Samuel and Charlie, I love you with all my being. I am so glad that you were a part of this journey. Spending time with you always made me forget I had had a tough day. You were surely the best "distractions" ever! Thank you.

Thank you all!

To God be the glory.

Colophon

Colophon

This work was supported by the European Commission (within the framework of the Erasmus-Mundus joint doctorate “EGS-ABG”), Breed4Food (a public-private partnership in the domain of animal breeding and genomics) and Hendrix Genetics.

The data used in this thesis were provided by Hendrix Genetics.

The cover of this thesis was designed by Kingsley Yaw Korsah and Enoch Osei Sarfo.

This thesis was printed by Digiforce | Proefschriftmaken.nl, De Limiet 26, 4131 NC, Vianen, the Netherlands.

ISSN: 1652-6880

ISBN (print version): 978-91-7760-530-0

ISBN (electronic version): 978-91-7760-531-7