Genotype by environment interactions in poultry breeding programs

Thinh Tuan Chu

Supervisor committee - Aarhus University

Main Supervisor

Dr. Elise Norberg
Senior researcher, Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark
Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

Co-supervisors

Prof. Dr. Just Jensen
Professor, Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark

Prof. Dr. Peer Berg
Senior researcher, Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark
Professor, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

Prof. Dr. Ir. Hans Komen
Professor, Animal Breeding and Genomics, Wageningen University & Research, 6709 PG Wageningen, the Netherlands

Dr. Ir. John W. M. Bastiaansen
Researcher, Animal Breeding and Genomics, Wageningen University & Research, 6709 PG Wageningen, the Netherlands

This research was conducted under the joint auspices of the Graduate School of Science and Technology (GSST), Aarhus University, Denmark and the Graduate School of Wageningen Institute of Animal Sciences (WIAS), Wageningen University & Research, the Netherlands as part of the Erasmus Joint Doctorate Program “EGS-ABG”.

Thesis committee - Wageningen University

Promotors

Prof. Dr. Ir. Hans Komen
Professor, Animal Breeding and Genomics, Wageningen University & Research, 6709 PG Wageningen, the Netherlands

Co-promotors

Dr. Ir. John W. M. Bastiaansen
Researcher, Animal Breeding and Genomics, Wageningen University & Research, 6709 PG Wageningen, the Netherlands

Other members

Dr. Elise Norberg, Aarhus University, Denmark
Prof. Dr. Just Jensen, Aarhus University, Denmark
Prof. Dr. Peer Berg, Aarhus University, Denmark

This research was conducted under the joint auspices of the Graduate School of Science and Technology (GSST), Aarhus University, Denmark and the Graduate School of Wageningen Institute of Animal Sciences (WIAS), Wageningen University & Research, the Netherlands as part of the Erasmus Joint Doctorate Program “EGS-ABG”.

Genotype by environment interactions in poultry breeding programs

Improving genetic gain of breeding programs for poultry reared under commercial production conditions

Thinh Tuan Chu

Thesis

submitted in fulfillment of the requirements for the joint degree of doctor between
Aarhus University
by the authority of the Head of Graduate School of Science and Technology, and
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 Head of Graduate School of Science at Aarhus University and by the Academic Board of Wageningen University
to be defended in public
on Friday November 8, 2019
at 9.30 a.m. in Foulum, Aarhus University.

Thinh Tuan Chu
Genotype by environment interactions in poultry breeding programs

PhD thesis, Aarhus University, Denmark and Wageningen University, the Netherlands (2019)

With summaries in English and Danish

ISBN: 978-87-93787-80-3

ISBN: 978-94-6395-209-5
DOI: https://doi.org/10.18174/506477

Abstract

Environmental differences between the breeding (B) and commercial production (C) environments may lead to genotype-by-environment interactions (GxE) i.e. re-ranking of breeding values of animals in the two environments. A substantial re-ranking implies genetic progress achieved in breeding programs is not realized in performance of production animals. The issues of GxE are not new and several solutions exist, however, there has not been much focus on solutions for breeding programs for poultry. This PhD-project investigated GxE interactions in breeding programs for poultry and solutions to improve genetic progress in these breeding programs.

A strong GxE interaction for body weight (BW) traits was found in broilers that were raised in B and C environments. Indications of GxE were significant re-ranking of breeding values, heterogeneous variances and different heritability for BW under B and C conditions. The genetic correlations between BW traits measured in B and C environments were in the range 0.48-0.54. Genetic variances of C traits were more than 2 times higher than those of B traits. Heritability of C traits (0.31-0.37) were higher than those of B traits (0.27-0.30).

In this thesis, several approaches to improve genetic gains of the poultry breeding programs in the presence of GxE have been investigated: phenotyping strategies, optimal modelling of traits, use of group records, and the use of genomic information. Different phenotyping strategies were compared in a breeding program for broilers that used genomic selection. It was found that when the genetic correlations between traits measured in B and C were 0.5 and 0.7, allocation of 70% and 30% hatched birds to B and C environments, respectively, for phenotype testing led to the highest genetic gains among the compared phenotyping strategies. When the genetic correlation was 0.9, moving birds to C did not improve genetic gains of the breeding scheme due to reduced selection intensity. Increasing proportion of birds moved to C (from 15 to 45%) could reduce rate of inbreeding of the breeding program.

Optimal modelling of traits was explored in a genetic analysis that was carried out for BW in broilers at different ages raised in a commercial environment. A statistical model was developed with the aim to increase predictive ability of the model for the traits affected by maternal effects. A criterion for the development of the statistical model was based on correlation between EBVs and corrected phenotypes of half-sib individuals. The statistical model also accounted for heterogeneous variances between sexes.

In breeding programs for village chicken, where strong GxE interactions are expected, the use of group records was a good option to increase genetic gain of the breeding programs. The use of group records

from villages significantly improved genetic gains compared to the scheme without birds tested in the village although group records led to a slightly lower genetic gain compared to individual records.

In addition, the use of genomic information was exploited to improve genetic gain of poultry breeding programs in the presence of GxE. Compared to pedigree, genomic information increased accuracy of the prediction from individual records. The use of combined pedigree and genomic information in the ssGBLUP prediction from individual records substantially increased accuracy of EBVs of C traits by 31-37%, and reduced bias of prediction for genotyped selection candidates. Genomic information was also utilized to form groups, so that accuracy of the prediction from group records increased compared to the use of pedigree information.

Overall, differences between the breeding and production environments can lead to substantial GxE interactions. In the presence of GxE interactions, a breeding program for poultry should establish recording systems under the production environments in either individual or group records in order to ensure maximum genetic gains and provide customers with genotypes well adapted to the production environments. In addition, an optimal cross-validation procedure for the choice of statistical models is needed for genetic evaluations in poultry breeding programs as better modelling of traits is a low-cost approach to improve accuracy of selection.

In memory of my grandmother, who gave me her best 

Contents

1

General Introduction

Poultry species play a key role in animal industry. The world production in 2017 was 116 million tons of poultry meat (chicken, duck, goose and guinea fowl) and 87 million tons of eggs (hen and other birds), compared to 120 million tons of pig meat and 66 million tons of beef meat (FAO, 2017). Of the world production, chicken accounted for 94% of the poultry meat and 92% of the poultry eggs. Poultry products including meat and eggs have little religious or cultural constrains associated with human consumption. Per capita consumption has increased greatly across both developed and underdeveloped nations (FAO, 2017). Global prices of poultry products are relatively lower than other animal protein products like pork, beef and dairy products (Tavárez and Solis de los Santos, 2016). The relatively lower price is credited to remarkable productivity and efficiency of poultry production. This has been achieved through improvements in genetics, nutrition and management, of which genetics and breeding account for approximately 80-90% within the last 50 years (Havenstein *et al.*, 2003).

In major parts of the world, commercial breeds of broiler chicken are used widely in commercial intensive systems. The organizational structure of poultry production chains follows a pyramidal structure. A relatively small breeding population of pure lines is on the top of the pyramid. The end-products of commercial chicken, often 4-way crossbred, are at the base of the pyramid with very large populations. It has been reported that the progress of genetic merit of the end-product chicken is not equivalent to the progress of genetic gain of the pure lines. Several possible explanations exist, including the genetic lag time from pure lines to end-products, crossbreeding, correlation between purebred and crossbreed performance and genotype-by-environment interaction (GxE) effects. Differences between breeding and production environments may lead to GxE interactions.

In some parts of the world, indigenous breeds of poultry are preferred in low-input village production systems. Village smallholders are reluctant to the introduction of exotic breeds or high-yielding hybrids because the application of exotic breeds requires high input and it brings in relatively lower economic returns than indigenous poultry under low input production systems (FAO, 2010; Okeno *et al.*, 2013). To improve livelihood of the poor people in the rural and peri-urban regions of Sub-Saharan Africa, there are breeding programs for indigenous chicken such as the breeding program initiated in 2008 at Debre Zeit Agricultural Research Centre, Ethiopia, for the dual-purpose Horro chicken (Dana *et al.*, 2011). For ease of implementation, the programs are often carried out at research stations to improve productivity of indigenous chicken. Smallholders in villages are targeted users of the improved indigenous breeds. However, rearing conditions are different between station and village environments, i.e. GxE interactions can be expected.

1.1. Breeding environments versus production environments

Majority of commercial broiler chicken raised worldwide are derived from broiler breeding stocks of large breeding companies such as Aviagen, Cobb-Vantress, and Hubbard (commercial breeding companies in alphabetical order) (Hiemstra and ten Napel, 2013). In these breeding companies, performance testing and selection are carried out in purebred lines, and major genetic gain are achieved in the breeding environment. A major difference between this breeding environment (B) and commercial production environment (C) is due to differences in hygienic conditions. To avoid the risk of losing the lines and to prevent a worldwide spread of diseases, a very strict bio-security are applied to the B environment, e.g. isolated farms, decontaminated feeds with regular bacteriological tests, stringent in-out policies, health monitoring of birds, strict logistic policies and systematical tests on a range of diseases with large government-certified laboratories (Hiemstra and ten Napel, 2013). In contrast, broilers are raised in less hygienic conditions of the C environment. With highly bio-secure procedures, purebred birds in the Aviagen commercial breeding company are completely disease-free from Salmonellae, Mycoplasma, Leucosis, Avian Influenza and Newcastle disease (Hiemstra and ten Napel, 2013). Meanwhile, these diseases remain chronic problems in many commercial poultry flocks (De Boeck *et al.*, 2015; EFSA, 2017; Shamim *et al.*, 2015). Kapell *et al.* (2012) shows that the incidence of food-pad dermatitis in birds of C environment was up to 3.5 times more than birds in B environment. The environmental difference between B and C may be also related to litter management. Litter, commonly wood shavings are used, is loose, dry, free-flowing in B environment (Hiemstra and ten Napel, 2013). At the end of each cycle, the litter is removed completely, housing was disinfected and floor is filled with fresh wood shavings (Kapell *et al.*, 2012). In some C conditions, at the end of each cycle, half of the litter was retained, mechanically conditioned, and then topped up with fresh wood shavings (Kapell *et al.*, 2012). In addition, feed compositions used in B environment may have higher level of proteins and metabolizable energy than the feed used in C environment (Kapell *et al.*, 2012). Another key factor is stress due to post-hatch handling of a-day-old chicks, particularly transportation procedures from hatchery to the commercial production farms. In B environment, a-day-old chicks are directly placed to the on-site rearing facility without transportation procedures, thus the stress on the chicks is minimum. In contrast, the chicks transferred to on-farm facilities are subjected to suboptimal conditions during transportation that may cause stress to a-day-old chicks in C. For example, Jacobs *et al.* (2017) found that a-day-old chicks transported for 1.5h showed a significantly lower level of stress than the chicks transported for 11h. Bergoug *et al.* (2013) showed that birds received no transportation had higher body weight (BW) and lower incidence of severe

foot-pad dermatitis than birds received 4h and 10h of transportation from hatchery to farms. Such differences between B and C environments may lead to GxE interaction.

In breeding programs for indigenous chicken for low-input village production systems, the B environment is at the research station where the birds have *ad libitum* access to nutritionally adequate feed and water, semi-controlled environments for light and ventilation, well-protected housing, deep litter floor and proper vaccination (Dana, 2011; Wondmeneh, 2015). In contrast, the C environment for the indigenous chicken in village is typically a low input system. In the village conditions, the birds are subjected to a combination of low food availability, sub-optimal diet, prevalence of diseases, predators and other social interaction factors. A high proportion of feed is from scavenging activities, and supplementary feed have poor quality (Dana, 2011). The birds are exposed to different diseases with limited or no vaccination provided (Dana, 2011). The flock size of chicken is only 7-20 from a household, and the birds are raised together with other livestock species as well (Dana, 2011). The differences between B and C environments, in the situation for indigenous chicken, are more severe than those for commercial broiler breeds, and thus a high level of GxE interaction is expected.

1.2. GxE interaction and its consequences on breeding programs

As variability is a rule of nature, different expression of genotypes over altered environments, known as GxE interactions, is expected (Mathur, 2003). However, the GxE interactions concern breeding programs only when re-ranking of genotypes in different environments occurs. Figure 1.1 shows different types of GxE interactions. There are changes in trait expression of genotypes in different environments for types III, but the changes do not result in genetic re-ranking of genotypes in the two environments. For types IV, V and VI, the difference in environments affects expression of traits and change the genetic ranking of animals. In a breeding program, the change in genetic ranking of animals means that the best animals selected based on performance in one environment are not the best in the target environment. If the re-ranking is substantially large between B and C environments, the breeding program carried out in B environment is not efficient or beneficial.

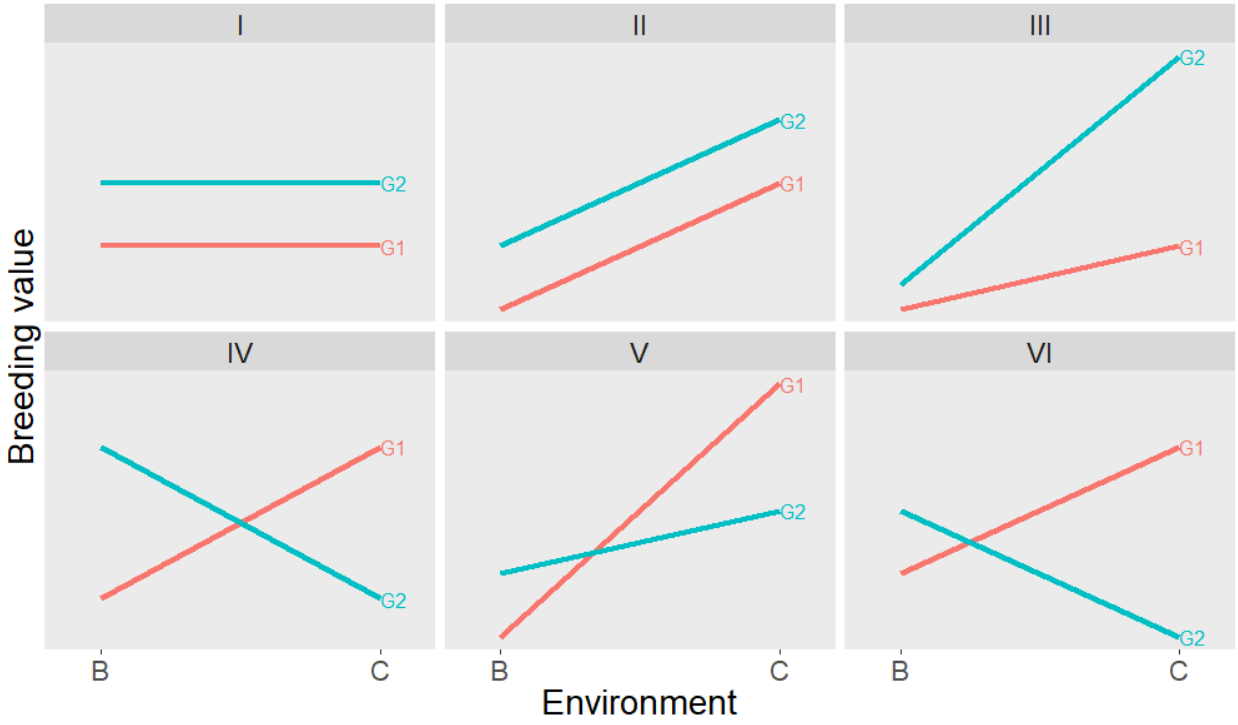


Figure 1.1. Types of GxE interactions: No GxE interactions (Type I and II); GxE interactions without re-ranking (Types III); GxE interactions with re-ranking (Types IV, V and VI). G1 and G2 are genotypes; B and C are two environments.

To model GxE, traits expressed in two environments can be defined as two correlated traits. Indications of GxE can be different performances, genetic correlation lower than 1, heterogeneous variances and different heritability between the traits measured in the two environments. Of these indications, the genetic correlation between traits measured in two environments represents the magnitude of GxE interaction as it indicates the level of re-rankings between animals. Indications of GxE in poultry have been found for breeding programs due to the differences between breeding and commercial production environments for commercial breeds (Kapell *et al.*, 2012; Long *et al.*, 2008; N'Dri *et al.*, 2007; Ye *et al.*, 2006) and between on-station and on-farm environments for indigenous breeds (Bekele *et al.*, 2009; Lwelamira, 2012). G×E interactions for different traits were found: bodyweight (r_g of 0.46-0.82), foot-pad dermatitis (r_g of 0.78-0.82) and meat quality (r_g of 0.84-0.93) (Kapell *et al.*, 2012; Lwelamira, 2012; N'Dri *et al.*, 2007).

1.3. GxE interaction and selection

Selection that is carried out under conditions which is most favourable for the expression of the genotype has been adopted widely in breeding programs because of ease of implementation. However, when GxE interactions are substantial, this approach is not efficient. In addition, to expand the market widely and internationally, breeding companies need to provide their customers with genotypes well adapted to the C environment on specific markets. Long-term selection of animals in the favourable environment may increase environmental sensitivity of the animals (Falconer and Mackay, 1996; van der Waaij, 2004). Increased environmental sensitivity could result in poor performance, leg problems and health issues when the animals are placed in unfavourable environments (Deeb and Cahaner, 2002; Deeb *et al.*, 2002; Falconer and Mackay, 1996; Kapell *et al.*, 2012; Kolmodin *et al.*, 2003; van der Waaij, 2004). The poor performance and health issues can have negative effects on economic returns and animal welfare.

To deal with GxE interactions, one can adjust the production conditions to make them similar to the conditions where animals are selected, or one can phenotype testing all test selection candidates under the production environments. Both are difficult from an economically and logically point of view. A compromise between these two options is to keep selection candidates under the selection environment and record phenotypes of relatives of the candidates in the production environment. The relatives can be sibs, progeny, or descendants of the selection candidates. This approach is being carried out in breeding programs for purebred broilers where full-sibs and half-sibs of the selection candidates are reared in farms replicating a range of commercial-like conditions (Kapell *et al.*, 2012).

Because the end-product of commercial animals is often 3 or 4-way crossbred in the pyramidal structure, the other approach is to record phenotypes of the crossbred in C environments. This approach is used in pig breeding programs that account for GxE interactions and genotype by genotype interactions (Dekkers, 2007; Wientjes and Calus, 2017). This approach accounts for the difference in genetic backgrounds between purebred and crossbred animals. However, this approach may be difficult for poultry breeding programs because recording of pedigree for crossbred birds in C environments is challenging. In addition, when selection of purebred is based on crossbred performance, accuracy of selection may be low because the relationship coefficients between candidates and crossbred animals are often low.

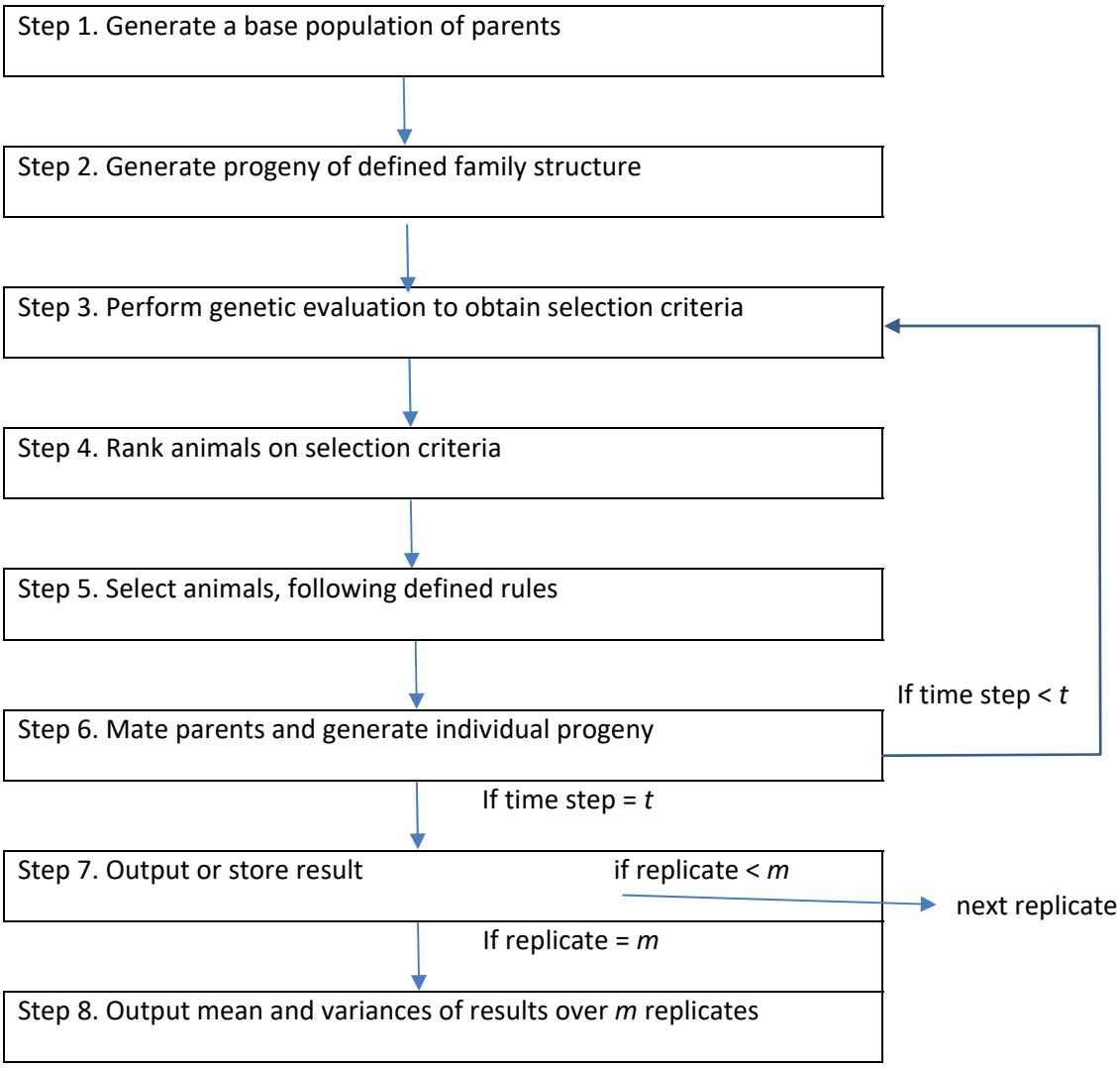
1.4. Methods to design breeding programs

Simulation is a good tool for designing and evaluating different breeding schemes. Compared to large-scale field-testing of a breeding scheme, simulation is a comparatively fast and inexpensive approach.

Simulation can be used to test numerous strategies for genotyping, phenotyping, breeding structure and criteria for selection, which are “too luxury” for a field test. To deal with GxE, we need to adjust the conventional breeding program doing phenotype-testing under B conditions only. Simulation has been used to design breeding programs in the presence of GxE interaction by others (Dekkers, 2007; Mulder and Bijma, 2005; van Grevenhof and van der Werf, 2015). From a methodological point of view, simulation can be carried out stochastically and deterministically to quantify the expected genetic gain and the expected rate of inbreeding for breeding programs (Dekkers, 2003).

In a stochastic simulation, an entire population of “real” animals is simulated. Stochastic simulation models of breeding programs generally follow the scheme as shown in Figure 1.2. However, polygenic and genomic models are different in how the base population is created and how breeding values are simulated and inherited. With a polygenic model, breeding values and phenotypes of each individual in the base population are generated by random sampling from pre-defined distributions. Breeding values of progeny are generated based on parents’ breeding values and the Mendelian sampling contribution. With a genomic model, the genome (QTL and markers) of each individual in the base population is created through number of historical generations to establish linkage equilibrium for QTL and markers. Inheritance and recombination of QTL and markers from parents to descendants follows the standard principles of Mendelian inheritance. Breeding values of each individual are generated based on QTL effects. The QTL effects are sampled from distributions, and then scaled to achieve the pre-defined distributions of breeding values of individuals in the base population.

Figure 1.2. General schematic of a stochastic simulation of a breeding program with *t* time steps of *m* replicates (adapted from Dekkers (2003)).



An important advantage of stochastic simulation is that stochastic simulation can mimic a complex breeding program in specific desired details because the individual animal is simulated (Dekkers, 2003). Therefore, stochastic simulation can be very flexible and precise, and it can be used to validate results from deterministic simulation (Dekkers, 2003). Because stochastic simulation is based on random sampling, many replicates are required to obtain the mean expected response. This replication process may be time-consuming and computation-expensive, but it is also an advantage because the variance of the response can be estimated. Compared to deterministic simulation, the main disadvantage of stochastic simulation is time-consuming and computation-expensive procedures, particularly for the genomic simulation models. There are several software programs and packages available for stochastic

simulation of animal breeding programs such as ADAM (Pedersen *et al.*, 2009), DCBSP (Medrano *et al.*, 2010) and AlphaSim (Faux *et al.*, 2016). Among these, ADAM, which has been intensively developed for more than 10 years, is a relatively comprehensive software for simulation of animal and plant breeding programs. Compared to other simulation programs, ADAM enables us to simulate numerous features for very complex breeding programs: polygenic (infinitesimal) model, genomic (finite locus) model, overlapping breeding cycles, different statistical models, storages of sperm and embryos, and different strategies for phenotyping, genotyping, selection, mating and family structures (Liu *et al.*, 2019; Pedersen *et al.*, 2009). With the infinitesimal simulation models, stochastic simulation of a complex breeding program is relatively fast with ADAM (Liu *et al.*, 2019; Pedersen *et al.*, 2009).

Deterministic simulation is a quick method to quantify the expected response (genetic gain and inbreeding) from alternative breeding schemes. The advantage of the deterministic simulation is much less computation. The deterministic method does not simulate the breeding program on the individual animal level, but derive (deterministic) equations to predict the expected response. However, the disadvantage of the deterministic method is that the method cannot precisely model very complex breeding programs (Dekkers, 2003). It is relatively difficult to write or code for deterministic simulation of breeding programs. There are only a few software programs available for deterministic simulation compared to stochastic simulation programs because each breeding program requires a different set of equations to predict the expected response. Software SelAction is a deterministic simulation tool for designing breeding programs (Rutten *et al.*, 2002).

1.5. Genomic information

In dealing with GxE in a broiler breeding program, the classical method is to test sibs in B and C environments and use pedigree-based BLUP for selection. However, given a breeding program with limited testing capacity and bio-security restriction, selection based on the classical method has low accuracy because sib-testing using pedigree-based BLUP has low prediction accuracy of the Mendelian sampling term. In the era of genomic selection, it is expected that high density of markers can help to improve the low accuracy, thus improve genetic gains of breeding programs in the presence of GxE. Compared to the use of pedigree information only, genomic information has shown its benefits for increased accuracy of selection in simulations (Andonov *et al.*, 2017; Christensen and Lund, 2010; Hayes *et al.*, 2009; Lourenco *et al.*, 2013; Meuwissen *et al.*, 2001; Putz *et al.*, 2018) and empirical studies of chicken (Alemu *et al.*, 2016; Chen *et al.*, 2011a; Chen *et al.*, 2011b; Momen *et al.*, 2017; Wolc *et al.*, 2011),

cattle (Gao *et al.*, 2018; Lee *et al.*, 2017; Li *et al.*, 2016; Lourenco *et al.*, 2015) and pig (Christensen *et al.*, 2012; Guo *et al.*, 2015; Putz *et al.*, 2018; Xiang *et al.*, 2016) breeding programs.

However, to my knowledge, the benefits of genomic information in a breeding program for broiler chicken with GxE sib-testing designs has not been reported in empirical studies. Such a breeding program typically uses a selective genotyping strategy and has no phenotypes measured in C for selection candidates in B. Economically important traits in broilers such as BW and feed efficiency can be obtained before the birds are sexually mature. When resources for genotyping are limited, genotyping is applied to only potential parents with best performances for the traits. Compared to the random genotyping strategy, the selective genotyping strategy leads to higher genetic gain (Boligon *et al.*, 2012). However, the selective genotyping leads to an overestimation of genetic variances with single step GBLUP (ssGBLUP) (Cesarani *et al.*, 2019). With GxE sib-testing designs of the breeding program for broiler, selection candidates have no performances in C environment. The cross-validation of evaluation models that is based on the correlation between corrected phenotypes and EBVs cannot be used to compute accuracy of EBVs.

1.6. Traits affected by maternal effects

Body weight of chicken at the age that selection is carried out may be affected by the dam. Dams can affect chicken through several ways such as egg weight, nutrient contents of egg and levels of maternal antibodies transferred to chicken. For example, age of dam had significant effects on egg weight, egg quality and chick quality (Lapao *et al.*, 1999; Tona *et al.*, 2004). Genotypes of dams and/or exposures of dams to different specific diseases can result in different levels of maternal antibodies transferred from the dam to eggs and chicks (De Boeck *et al.*, 2015; Hamal *et al.*, 2006; Ismiraj *et al.*, 2019). Other studies (Tahir *et al.*, 2011; Tona *et al.*, 2004; Wolanski *et al.*, 2007) showed that dam could have effects on BW performance of broilers. To model BW of chicken in genetic analysis, different maternal factors have been included in the statistical models: age of dam, specific environmental herd of the dam, maternal additive genetic effect, covariance between maternal additive genetic effect and direct additive genetic effect, and permanent environmental maternal effect (Jasouri *et al.*, 2017; Koerhuis and Thompson, 1997; Maniatis *et al.*, 2013).

Failing to account for the maternal factors in the statistical model can lead to reduced accuracy of selection and increased bias of EBV prediction. For instance, when the permanent environmental maternal effect was present, but ignored in the prediction model, direct additive genetic variance was overestimated, accuracy of prediction was reduced and bias of prediction was increased (Su *et al.*, 2018).

Ignoring the maternal additive genetic effect and/or the permanent environmental maternal effect can lead to an overestimation of the direct additive genetic variance (Jasouri *et al.*, 2017). Failing to account for the covariance in the model can lead to a possible underestimation of the direct genetic and maternal genetic variances if the covariance is negative, or overestimation of the variances if the covariance is positive (Chapuis *et al.*, 1996).

Ignoring effects in modelling of a trait of interests can have negative consequence on the predictive ability of the model. However, the true model for field data is usually unknown in practice, thus the question is how we can find an optimal model for the trait of interest, particularly for the trait affected by maternal effects. For the development of a statistical model, animal breeders usually use cross-validation procedures to assess predictive ability or accuracy of EBV between competing models. The common approach of cross-validation is based on the correlation between corrected phenotypes and EBV of selection candidates as used in Christensen *et al.* (2012). For the trait affected by maternal effect, this approach may lead to a wrong conclusion on model comparisons. Information for the prediction of EBVs of selection candidates is largely from their full-sibs in the genetic evaluation of a breeding program for chicken. If the maternal effects are not accounted for appropriately in the model, the effects shared among full-sibs may influence the EBVs of the selection candidates. The correlation between corrected phenotypes and EBV would be overestimated for the statistical model that ignores the presence of maternal effects. Therefore, for optimal modelling of chicken BW, we may need a different cross-validation approach to assess competing models.

1.7. Group records

Prediction of EBVs from group records was shown to be feasible with pedigree-based BLUP in simulation studies (Olson *et al.*, 2006; Peeters *et al.*, 2013; Su *et al.*, 2018). However, to our knowledge, the use of group records has been limited in breeding programs. This may be because the additional accuracy of selection from the use of group records may be relatively small for the breeding programs where data recording on individual basis is typically the norm for all selection candidates. This view may change when strong GxE interactions due to environmental differences between B and C exist. A large number of records measured in C environment are required to improve genetic gain of the breeding programs in the presence of strong GxE interactions (Mulder and Bijma, 2005). However, in C environment, particularly village production systems, parts or all of the data is often recorded on groups of animals (e.g. egg production). Therefore, the question is how to use data recorded on groups to improve the breeding

programs in the presence of GxE. Prediction from group records has only been done with pedigree-based BLUP models, thus this thesis extents the use of group records with genomic information.

1.8. Objectives

This study focuses on GxE interactions due to differences between breeding and production environments, and propose solutions to improve accuracy of selection for poultry breeding programs that are carried out for commercial intensive systems and low-input village production systems. Five specific objectives were investigated and the results are presented in five corresponding papers:

1. Identify GxE interactions due to environmental differences between breeding and commercial production environments for broilers, and exploit the use of genomic information to increase accuracy of predicted breeding values in the presence of GxE.
2. Design genomic selection breeding schemes for commercial broiler chicken in the presence of GxE.
3. Develop statistical models to improve predictive ability of predicted breeding values for broiler breeding programs.
4. Design and compare breeding schemes for village poultry production.
5. Optimize grouping methods based on genomic information to improve accuracy of prediction from group records.

References

Alemu, SW, Calus, MPL, Muir, WM, Peeters, K, Vereijken, A & Bijma, P 2016, 'Genomic prediction of survival time in a population of brown laying hens showing cannibalistic behavior', *Genetics Selection Evolution,* vol. 48, no. 1, pp. 68.

Andonov, S, Lourenco, DAL, Fragomeni, BO, Masuda, Y, Pocrnic, I, Tsuruta, S & Misztal, I 2017, 'Accuracy of breeding values in small genotyped populations using different sources of external information-A simulation study', *J Dairy Sci,* vol. 100, no. 1, pp. 395-401.

Bekele, F, Gjoen, HM, Kathle, J, Adnoy, T & Abebe, G 2009, 'Genotype X environment interaction in two breeds of chickens kept under two management systems in Southern Ethiopia', *Tropical Animal Health and Production,* vol. 41, no. 7, pp. 1101-14.

Bergoug, H, Guinebretière, M, Tong, Q, Roulston, N, Romanini, CEB, Exadaktylos, V, Berckmans, D, Garain, P, Demmers, TGM, McGonnell, IM, Bahr, C, Burel, C, Eterradossi, N & Michel, V 2013, 'Effect of transportation duration of 1-day-old chicks on postplacement production performances and pododermatitis of broilers up to slaughter age', *Poultry Science,* vol. 92, no. 12, pp. 3300-3309.

Boligon, AA, Long, N, Albuquerque, LG, Weigel, KA, Gianola, D & Rosa, GJM 2012, 'Comparison of selective genotyping strategies for prediction of breeding values in a population undergoing selection', *Journal of Animal Science,* vol. 90, no. 13, pp. 4716-4722.

Cesarani, A, Pocrnic, I, Macciotta, NPP, Fragomeni, BO, Misztal, I & Lourenco, DAL 2019, 'Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies', *Journal of Animal Breeding and Genetics,* vol. 136, no. 1, pp. 40-50.

Chapuis, H, Tixier-Boichard, M, Delabrosse, Y & Ducrocq, V 1996, 'Multivariate restricted maximum likelihood estimation of genetic parameters for production traits in three selected turkey strains', *Genetics Selection Evolution,* vol. 28, no. 3, pp. 299.

Chen, C, Misztal, I, Aguilar, I, Tsuruta, S, Aggrey, S, Wing, T & Muir, W 2011a, 'Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens', *Journal of animal science,* vol. 89, no. 1, pp. 23-28.

Chen, CY, Misztal, I, Aguilar, I, Legarra, A & Muir, WM 2011b, 'Effect of different genomic relationship matrices on accuracy and scale', *J Anim Sci,* vol. 89, no. 9, pp. 2673-9.

Christensen, OF & Lund, MS 2010, 'Genomic prediction when some animals are not genotyped', *Genetics Selection Evolution,* vol. 42, no. 1, pp. 2.

Christensen, OF, Madsen, P, Nielsen, B, Ostersen, T & Su, G 2012, 'Single-step methods for genomic evaluation in pigs', *Animal,* vol. 6, no. 10, pp. 1565-71.

Chu, TT, Alemu, SW, Norberg, E, Sørensen, AC, Henshall, J, Hawken, R & Jensen, J 2018, 'Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 52.

Dana, N 2011, 'Breeding programs for indigenous chicken in Ethiopia, Wageningen University, The Netherlands.

Dana, N, vander Waaij, EH & van Arendonk, JAM 2011, 'Genetic and phenotypic parameter estimates for body weights and egg production in Horro chicken of Ethiopia', *Tropical Animal Health and Production,* vol. 43, no. 1, pp. 21-28.

De Boeck, C, Kalmar, I, Dumont, A & Vanrompay, D 2015, 'Longitudinal monitoring for respiratory pathogens in broiler chickens reveals co-infection of Chlamydia psittaci and Ornithobacterium rhinotracheale', *J Med Microbiol,* vol. 64, no. Pt 5, pp. 565-74.

Deeb, N & Cahaner, A 2002, 'Genotype-by-environment interaction with broiler genotypes differing in growth rate. 3. Growth rate and water consumption of broiler progeny from weight-selected versus nonselected parents under normal and high ambient temperatures', *Poult Sci,* vol. 81, no. 3, pp. 293-301.

Deeb, N, Shlosberg, A & Cahaner, A 2002, 'Genotype-by-environment interaction with broiler genotypes differing in growth rate. 4. Association between responses to heat stress and to cold-induced ascites', *Poultry Science,* vol. 81, no. 10, pp. 1454-1462.

Dekkers, JC 2003, 'Design and economics of animal breeding strategies', *Notes for summer short course. Iowa State University*, no.

Dekkers, JCM 2007, 'Marker-assisted selection for commercial crossbred performance', *J Anim Sci,* vol. 85, no. 9, pp. 2104-2114.

EFSA, EFSA 2017, 'The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016', *EFSA Journal,* vol. 15, no. 12, pp. 5077.

Falconer, DS & Mackay, TF 1996, 'Introduction to quantitative genetics ', 4th/Ed. Pearson Prentice Hall, England,464

FAO 2010, 'Chicken genetic resources used in smallholder production systems and opportunities for their development'. *In* '*FAO Smallholder Poultry Production Paper No. 5*' (P. Sørensen, ed.), Rome.

FAO 2017, 'FAO statistical database', accessed: accessed in 05/02/2019,

Faux, AM, Gorjanc, G, Gaynor, RC, Battagin, M, Edwards, SM, Wilson, DL, Hearne, SJ, Gonen, S & Hickey, JM 2016, 'AlphaSim: Software for Breeding Program Simulation', *Plant Genome,* vol. 9, no. 3, pp.

Gao, H, Koivula, M, Jensen, J, Strandén, I, Madsen, P, Pitkänen, T, Aamand, GP & Mäntysaari, EA 2018, 'Short communication: Genomic prediction using different single-step methods in the Finnish red dairy cattle population', *Journal of Dairy Science,* vol. 101, no. 11, pp. 10082-10088.

Guo, X, Christensen, OF, Ostersen, T, Wang, Y, Lund, MS & Su, G 2015, 'Improving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method', *J Anim Sci,* vol. 93, no.

Hamal, KR, Burgess, SC, Pevzner, IY & Erf, GF 2006, 'Maternal Antibody Transfer from Dams to Their Egg Yolks, Egg Whites, and Chicks in Meat Lines of Chickens', *Poultry Science,* vol. 85, no. 8, pp. 1364-1372.

Havenstein, G, Ferket, P & Qureshi, M 2003, 'Carcass composition and yield of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets', *Poultry Science,* vol. 82, no. 1509-1518.

Hayes, BJ, Visscher, PM & Goddard, ME 2009, 'Increased accuracy of artificial selection by using the realized relationship matrix', *Genet Res,* vol. 91, no.

Hiemstra, SJ & ten Napel, J 2013, 'Study of the impact of genetic selection on the welfare of chickens bred and kept for meat production', Final report of a project commissioned by the European Commission (DG SANCO 2011/12254).

Ismiraj, MR, Arts, JAJ & Parmentier, HK 2019, 'Maternal Transfer of Natural (Auto-) Antibodies in Chickens', *Poultry Science*, no.

Jacobs, L, Delezie, E, Duchateau, L, Goethals, K, Ampe, B, Buyse, J & Tuyttens, FAM 2017, 'Impact of transportation duration on stress responses in day-old chicks from young and old breeders', *Research in Veterinary Science,* vol. 112, no. 172-176.

Jasouri, M, Zamani, P & Alijani, S 2017, 'Dominance genetic and maternal effects for genetic evaluation of egg production traits in dual-purpose chickens', *Br Poult Sci,* vol. 58, no. 5, pp. 498-505.

Kapell, DN, Hill, WG, Neeteson, AM, McAdam, J, Koerhuis, AN & Avendano, S 2012, 'Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments', *Poultry Science,* vol. 91, no. 3, pp. 565-74.

Koerhuis, AN & Thompson, R 1997, 'Models to estimate maternal effects for juvenile bodyweight in broiler chickens', *Genet Sel Evol,* vol. 29, no.

Kolmodin, R, Strandberg, E, Jorjani, H & Danell, B 2003, 'Selection in the presence of a genotype by environment interaction: response in environmental sensitivity', *Animal Science,* vol. 76, no. 3, pp. 375-385.

Lapao, C, Gama, LT & Soares, MC 1999, 'Effects of broiler breeder age and length of egg storage on albumen characteristics and hatchability', *Poultry Science,* vol. 78, no. 5, pp. 640-645.

Lee, J, Cheng, H, Garrick, D, Golden, B, Dekkers, J, Park, K, Lee, D & Fernando, R 2017, 'Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle', *Genetics Selection Evolution,* vol. 49, no. 1, pp. 2.

Li, X, Lund, MS, Zhang, Q, Costa, CN, Ducrocq, V & Su, G 2016, 'Short communication: Improving accuracy of predicting breeding values in Brazilian Holstein population by adding data from Nordic and French Holstein populations', *Journal of Dairy Science,* vol. 99, no. 6, pp. 4574-4579.

Liu, H, Tessema, BB, Jensen, J, Cericola, F, Andersen, JR & Sørensen, AC 2019, 'ADAM-Plant: A Software for Stochastic Simulations of Plant Breeding From Molecular to Phenotypic Level and From Simple Selection to Complex Speed Breeding Programs', *Frontiers in Plant Science,* vol. 9, no. 1926, pp.

Long, N, Gianola, D, Rosa, GJ, Weigel, KA & Avendano, S 2008, 'Marker-assisted assessment of genotype by environment interaction: a case study of single nucleotide polymorphism-mortality association in broilers in two hygiene environments', *J Anim Sci,* vol. 86, no. 12, pp. 3358-66.

Lourenco, DA, Misztal, I, Wang, H, Aguilar, I, Tsuruta, S & Bertrand, JK 2013, 'Prediction accuracy for a simulated maternally affected trait of beef cattle using different genomic evaluation models', *J Anim Sci,* vol. 91, no. 9, pp. 4090-8.

Lourenco, DA, Tsuruta, S, Fragomeni, BO, Masuda, Y, Aguilar, I, Legarra, A, Bertrand, JK, Amen, TS, Wang, L, Moser, DW & Misztal, I 2015, 'Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus', *J Anim Sci,* vol. 93, no. 6, pp. 2653-62.

Lwelamira, J 2012, 'Genotype-Environment (GXE) Interaction for Body Weights for Kuchi Chicken Ecotype of Tanzania Reared Under Intensive and Extensive Management', *Global Journal of Medical Research,* vol. 12, no. 5, pp.

Maniatis, G, Demiris, N, Kranis, A, Banos, G & Kominakis, A 2013, 'Model comparison and estimation of genetic parameters for body weight in commercial broilers', *Canadian journal of animal science,* vol. 93, no. 1, pp. 67-77.

Mathur, PK 2003, 'Genotype-environment interactions: problems associated with selection for increased production'. *In* '*Poultry Genetics, Breeding and Biotechnology*' (W. M. Muir and S. E. Aggrey, eds.), CAB International, pp. 83-99.

Medrano, JF, Ahmadi, A & Casellas, J 2010, 'Dairy Cattle Breeding Simulation Program: A simulation program to teach animal breeding principles and practices', *Journal of Dairy Science,* vol. 93, no. 6, pp. 2816-2826.

Meuwissen, THE, Hayes, BJ & Goddard, ME 2001, 'Prediction of total genetic value using genome-wide dense marker maps', *Genetics,* vol. 157, no. 4, pp. 1819-29.

Momen, M, Mehrgardi, AA, Sheikhy, A, Esmailizadeh, A, Fozi, MA, Kranis, A, Valente, BD, Rosa, GJM & Gianola, D 2017, 'A predictive assessment of genetic correlations between traits in chickens using markers', *Genetics Selection Evolution,* vol. 49, no. 1, pp. 16.

Mulder, H & Bijma, P 2005, 'Effects of genotype× environment interaction on genetic gain in breeding programs', *Journal of Animal Science,* vol. 83, no. 1, pp. 49-61.

N'Dri, AL, Sellier, N, Tixier-Boichard, M, Beaumont, C & Mignon-Grasteau, S 2007, 'Genotype by environment interactions in relation to growth traits in slow growing chickens', *Genetics Selection Evolution,* vol. 39, no. 5, pp. 513-528.

Okeno, TO, Magothe, TM, Kahi, AK & Peters, KJ 2013, 'Breeding objectives for indigenous chicken: model development and application to different production systems', *Tropical Animal Health and Production,* vol. 45, no. 1, pp. 193-203.

Olson, KM, Garrick, DJ & Enns, RM 2006, 'Predicting breeding values and accuracies from group in comparison to individual observations', *Journal of Animal Science,* vol. 84, no. 1, pp. 88-92.

Pedersen, L, Sørensen, A, Henryon, M, Ansari-Mahyari, S & Berg, P 2009, 'ADAM: A computer program to simulate selective breeding schemes for animals', *Livestock Science,* vol. 121, no. 2, pp. 343-344.

Peeters, K, Ellen, ED & Bijma, P 2013, 'Using pooled data to estimate variance components and breeding values for traits affected by social interactions', *Genetics Selection Evolution,* vol. 45, no.

Putz, AM, Tiezzi, F, Maltecca, C, Gray, KA & Knauer, MT 2018, 'A comparison of accuracy validation methods for genomic and pedigree-based predictions of swine litter size traits using Large White and simulated data', *J Anim Breed Genet,* vol. 135, no. 1, pp. 5-13.

Rutten, M, Bijma, P, Woolliams, J & Van Arendonk, J 2002, 'SelAction: Software to predict selection response and rate of inbreeding in livestock breeding programs', *Journal of Heredity,* vol. 93, no. 6, pp. 456-458.

Shamim, A, ul Hassan, M, Yousaf, A, Iqbal, MF, Zafar, MA, Siddique, RM & Abubakar, M 2015, 'Occurrence and identification of Emeria species in broiler rearing under traditional system', *Journal of Animal Science and Technology,* vol. 57, no. 41.

Su, G, Madsen, P, Nielsen, B, Ostersen, T, Shirali, M, Jensen, J & Christensen, OF 2018, 'Estimation of variance components and prediction of breeding values based on group records from varying group sizes', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 42.

Tahir, M, Cervantes, H, Farmer, CW, Shim, MY & Pesti, GM 2011, 'Broiler performance, hatching egg, and age relationships of progeny from standard and dwarf broiler dams', *Poultry Science,* vol. 90, no. 6, pp. 1364-1370.

Tavárez, MA & Solis de los Santos, F 2016, 'Impact of genetics and breeding on broiler production performance: a look into the past, present, and future of the industry', *Animal Frontiers,* vol. 6, no. 4, pp. 37-41.

Tona, K, Onagbesan, O, De Ketelaere, B, Decuypere, E & Bruggeman, V 2004, 'Effects of age of broiler breeders and egg storage on egg quality, hatchability, chick quality, chick weight, and chick posthatch growth to forty-two days', *Journal of Applied Poultry Research,* vol. 13, no. 1, pp. 10-18.

van der Waaij, EH 2004, 'A resource allocation model describing consequences of artificial selection under metabolic stress', *J Anim Sci,* vol. 82, no. 4, pp. 973-81.

van Grevenhof, IE & van der Werf, JH 2015, 'Design of reference populations for genomic selection in crossbreeding programs', *Genetics Selection Evolution,* vol. 47, no. 1, pp. 1-9.

Wientjes, YCJ & Calus, MPL 2017, 'BOARD INVITED REVIEW: The purebred-crossbred correlation in pigs: A review of theory, estimates, and implications1', *Journal of Animal Science,* vol. 95, no. 8, pp. 3467-3478.

Wolanski, NJ, Renema, RA, Robinson, FE, Carney, VL & Fancher, BI 2007, 'Relationships Among Egg Characteristics, Chick Measurements, and Early Growth Traits in Ten Broiler Breeder Strains', *Poultry Science,* vol. 86, no. 8, pp. 1784-1792.

Wolc, A, Stricker, C, Arango, J, Settar, P, Fulton, JE, O'Sullivan, NP, Preisinger, R, Habier, D, Fernando, R, Garrick, DJ, Lamont, SJ & Dekkers, JCM 2011, 'Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model', *Genetics Selection Evolution,* vol. 43, no. 1, pp. 5.

Wondmeneh, E 2015, 'Genetic improvement in indigenous chicken of Ethiopia, Wageningen University, Netherlands.

Xiang, T, Nielsen, B, Su, G, Legarra, A & Christensen, OF 2016, 'Application of single-step genomic evaluation for crossbred performance in pig', *Journal of animal science,* vol. 94, no. 3, pp. 936-948.

Ye, X, Avendano, S, Dekkers, JC & Lamont, SJ 2006, 'Association of twelve immune-related genes with performance of three broiler lines in two different hygiene environments', *Poult Sci,* vol. 85, no. 9, pp. 1555-69.

2

Use of genomic information to exploit genotype by environment interactions for body weight of broiler chicken in bio-secure and production environments

Thinh T. Chu[1,2*], John W. M. Bastiaansen[2], Peer Berg[1,3], Helene J. S. Rome[1], Danye Marois[4], John Henshall[4], Just Jensen[1]

[1] Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark

[2] Wageningen University & Research, Animal Breeding and Genomics, 6709 PG Wageningen, The Netherlands

[3] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[4] Cobb-Vantress Inc., Siloam Springs, Arkansas 72761-1030, United States of America

* Corresponding author

Published in *Genetics Selection Evolution*, 2019, 51:50

https://doi.org/10.1186/s12711-019-0493-3

Abstract

Background

The increase in accuracy of prediction by using genomic information has been well-documented. However, benefits of genomic information and methodology for evaluation are missing when genotype by environment interactions (GxE) exist between bio-secure breeding (B) environments and commercial production (C) environments. This study explored (1) GxE interactions for broiler body weight (BW) at week 5 and 6, and (2) benefits of using genomic information for prediction of BW traits when selection candidates were raised and tested in B environment and close relatives were tested in C environment.

Methods

A pedigree-based BLUP multivariate model was used to estimate variance components and predict breeding values (EBV) of BW traits at week 5 and 6 measured in B and C environments. A ssGBLUP model that combined pedigree and genomic information was used to predict EBV. Cross-validations were based on statistics of correlation, mean difference and regression slope for EBV that were estimated from full and reduced datasets. Those statistics were indicators of population accuracy, bias and dispersion of EBV prediction for EBV of B and C traits. Validation animals were genotyped and non-genotyped birds in B environment only.

Results

Several indications of GxE interactions due to environmental differences were found for BW traits including significant re-ranking, heterogeneous variances and different heritability for BW measured in B and C environments. The genetic correlations between BW traits measured in B and C environments were in the range 0.48-0.54. The use of combined pedigree and genomic information increased population accuracy of EBV, and reduced bias of EBV prediction for genotyped birds compared to the use of pedigree information only. A slight increase in accuracy of EBV also occurred for non-genotyped birds, but bias of EBV prediction increased for non-genotyped birds.

Conclusions

A strong GxE interaction is found for BW traits of broilers measured in B and C environments. The use of combined pedigree and genomic information substantially increases population accuracy of EBV for genotyped birds in B compared to the use of pedigree only.

Keywords: GxE; genomic selection; body weight; broiler.

2.1. Introduction

The difference in production conditions between highly bio-secure breeding (B) and commercial production environments (C) can lead to genotype by environment interaction (GxE) in broiler chicken. Indications of GxE may include heterogeneous variances, different heritability and correlation of less than unity between the same trait expressed under B and C conditions. To model GxE, the same trait expressed in the two environments can be defined as two correlated traits. Identification of the presence of GxE, especially genetic correlation between the B and C traits, is important in optimizing breeding programs. For example, Chu *et al.* (2018) showed that the genetic correlation between traits measured in B and C can change the optimal proportion of birds to be tested in B versus C environments. For body weight (BW) of broiler chicken, GxE interactions have been found with genetic correlations of 0.46-0.69 (Kapell *et al.*, 2012), 0.74-0.76 (N'Dri *et al.*, 2007) and 0.75-0.76 (Lwelamira, 2012) between traits measured in B and C. The ultimate goal of a breeding program for broilers is genetic gains of birds' performance in C environment only. To improve the genetic gains, sib-testing of purebred birds in B and C environments is an option (Kapell *et al.*, 2012). Due to bio-security restrictions, only birds in B are selection candidates, and birds in C provide information on C performance only. Because of limited reproductive capacity of broiler dams, a restricted number of birds can be moved to C for phenotype testing, and thus accuracy of prediction for performance in C might be relatively low with pedigree-based BLUP prediction (PBLUP) (Chu *et al.*, 2018). In this situation, genomic information can be of interest to improve accuracy of prediction.

Genomic selection has captured growing interests in poultry breeding programs because of the higher accuracy of prediction compared to pedigree-based evaluation. The increase in accuracy of prediction from using dense genotypes is due to better measuring the relationships between animals and a better prediction of the Mendelian sampling terms (Hayes *et al.*, 2009). The better explanation of relationships may improve the movement of information from birds in C to selection candidates in B. The benefit of genomic selection over pedigree-based selection has been well documented in simulations (Andonov *et al.*, 2017; Christensen and Lund, 2010; Hayes *et al.*, 2009; Lourenco *et al.*, 2013; Meuwissen *et al.*, 2001; Putz *et al.*, 2018) and empirical studies of chicken (Alemu *et al.*, 2016; Chen *et al.*, 2011a; Chen *et al.*, 2011b; Momen *et al.*, 2017; Wolc *et al.*, 2011), cattle (Gao *et al.*, 2018; Lee *et al.*, 2017; Li *et al.*, 2016; Lourenco *et al.*, 2015) and pig (Christensen *et al.*, 2012; Guo *et al.*, 2015; Putz *et al.*, 2018; Xiang *et al.*, 2016) breeding schemes. However, none of the empirical studies has reported the benefit of genomic information in a breeding program with GxE sib-testing designs. In such a program, a multivariate joint model is required to model traits measured in commercial and breeding environments. When the number of genotyped individuals and SNPs are numerous, the use of a multi-trait model can be computationally challenging, and estimation of variance components from the model that uses a realized genomic

relationship matrix could be tedious. Model-based accuracy or individual accuracy of EBV cannot be computed because obtaining prediction error variances by direct inversion of the left hand side of mixed model equations is infeasible. Cross-validation strategies for accuracy of EBV based on correlation between estimated breeding values (EBV) and corrected phenotypes cannot be applied. In GxE sib-testing breeding program, validation animals are birds in B, and these birds do not have corrected phenotypes of C traits. Legarra and Reverter (2018) proposed cross-validation measures that can compare competing prediction models in situations of breeding programs where traits are influenced by GxE. These validation measures (Legarra and Reverter, 2018) are based on statistics of EBV estimated from full and reduced datasets.

In addition, a typical breeding program utilizing genomic selection in broiler chicken often applies a selective genotyping strategy. In broilers, important traits like BW and feed efficiency can be obtained before sexual maturity. In situations of limited resources for genotyping, only a proportion of birds that are potential parents with best performances will be genotyped. The selective genotyping strategy can increase accuracy of selection and genetic gain compared to random genotyping strategies (Boligon *et al.*, 2012). However, the selective genotyping can create bias and lead to overestimation of genetic variances when ssGBLUP is employed to estimate variance components (Cesarani *et al.*, 2019). To utilize genomic information, as well as pedigree and phenotypes of non-genotyped birds in genetic evaluation, ssGBLUP models (Christensen *et al.*, 2012) can be used. However, accurate prediction of breeding values requires accurate estimates of variance components. An animal model using the pedigree relationship matrix is currently recommended for estimation of variance components in the situation of selective genotyping (Cesarani *et al.*, 2019).

Two main objectives of our study were: (1) to explore genotype by environment interaction for body weight (BW) in broilers raised in breeding bio-secure (B) and commercial production (C) environments, and (2) to use genomic information to increase accuracy of predicted breeding values of birds in B for BW traits measured in C environment.

2.2. Methods

Data

Data obtained from the poultry breeding company, Cobb-Vantress, included body-weight (BW) performances of purebred broiler chicken tested in breeding bio-secure (B) and standard commercial production (C) environments. The data had BW records from 16 time steps (TS) of selection that covered roughly 2.5 generations. Birds hatched at each TS were transferred to either B or C environment for

phenotype testing. Each full sib group was split between B and C so that each bird would have full and half sibs in both environments. Parents were selected from birds tested in B only. In other words, parents did not have any phenotypic records in C environments. In each TS, all offspring birds in C were hatched at the same time while offspring birds in B were hatched at several successive time points. Sires and dams of offspring birds in each TS were from several previous TS.

The BW of the broilers tested in B were recorded once, at 6 weeks of age (BW6.B), for TS 1-10 and recorded once or twice, at 5 and 6 weeks of age (BW5.B and BW6.B), for TS 11-16. All birds in B at TS 11-16 had BW5 records, but only 33% of those birds had BW6 records. The BW of the broilers tested in C were recorded at 5 and 6 weeks of age (BW5.C and BW6.C) for TS 5-10 and only at 5 weeks of age (BW5.C) for TS 11-16. The same data editing as in Chu *et al.* (2019) was carried out with removal of records for birds with unidentified sex, missing factors or duplicated records. Records of BW that were beyond four standard deviation units from the mean were also removed for each of the four BW record types. In total, 0.04% of all BW records were removed. After data editing, the number of birds in B and C were 54,757 and 15,412, respectively, with total of 61,589 and 23,569 BW records, respectively. The birds with BW records were from 319 sires and 1,528 dams. The pedigree covered roughly 3.5 generations back from the youngest birds and comprised 70,174 birds.

A medium density SNP chip with 55,792 markers was used for genotyping (Illumina, San Diego, CA, USA). Quality control was carried out that set missing rate for SNPs at <0.05 and call rate for birds at >0.95. Also, SNP markers with minor allele frequency of <0.01 were removed. After quality control, 39,767 genotyped birds with 50,562 SNPs remained for constructing the genomic relationship matrix. All parents had genotype information. Although all birds in C were genotyped, after quality control, genotyping information of a few birds in C was not used in constructing genomic relationship matrix. Only a proportion of the birds in B were genotyped.

Statistical models

From preliminary results, male and female BW had different variances, but high correlations that reflects scaling effects (Chu *et al.*, 2019). Modelling male and female BW as two traits led to convergence problems because of parameters at the edge of the parameter space. To model heterogeneous variances between male and female BW, standardization was applied to male and female BW separately (Chu *et al.*, 2019). Male and female phenotypic records of the four BW traits were standardized to their corresponding phenotypic standard deviations that were estimated in the following univariate model:

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Za}+\mathbf{Wc}+\mathbf{e} \qquad (2.1)$$

where **y** is a vector of male or female phenotypic records of BW at normal scale; **b** is a vector of fixed factors of hatch TS of the bird, TS of the parents and dam age in classes of one week. Matrices of **X**, **Z**, and **W** are incidence matrices. Vectors **a**, **c** and **e** are the direct additive genetic effect, permanent environmental maternal effect and residual, respectively. These random effects were assumed to be normally distributed: $\mathbf{a} \sim \mathbf{N}[\mathbf{0}, \mathbf{A}\sigma_a^2]$, $\mathbf{c} \sim \mathbf{N}[\mathbf{0}, \mathbf{I_d}\sigma_c^2]$ and $\mathbf{e} \sim \mathbf{N}[\mathbf{0}, \mathbf{I}\sigma_e^2]$, where **A** is the pedigree relationship matrix; $\mathbf{I_d}$ is the identity matrix for dams; **I** is the identity matrix for individual birds; σ_a^2, σ_c^2 and σ_e^2 are variances at normal scale of BW.

Standardization was applied, so that each phenotypic record was divided by the corresponding phenotypic standard deviation estimated from model (2.1). A multi-trait PBLUP model was used to estimate variance components and predict EBV from the standardized phenotypic records of BW in B and C environments. The four-trait PBLUP model (2.2) with heterogeneous residual variance for sexes was as follows:

$$\begin{bmatrix} \mathbf{y_{5B}^{m0}} \\ \mathbf{y_{5B}^{f0}} \end{bmatrix} = \begin{bmatrix} \mathbf{X_{5B}^{m}} & 0 \\ 0 & \mathbf{X_{5B}^{f}} \end{bmatrix} \begin{bmatrix} \mathbf{b_{5B}^{m}} \\ \mathbf{b_{5B}^{f}} \end{bmatrix} + \begin{bmatrix} \mathbf{Z_{5B}^{m}} & 0 \\ 0 & \mathbf{Z_{5B}^{f}} \end{bmatrix} \mathbf{a_{5B}} + \begin{bmatrix} \mathbf{W_{5B}^{m}} & 0 \\ 0 & \mathbf{W_{5B}^{f}} \end{bmatrix} \mathbf{c_{5B}} + \begin{bmatrix} \mathbf{e_{5B}^{m}} \\ \mathbf{e_{5B}^{f}} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{y_{6B}^{m0}} \\ \mathbf{y_{6B}^{f0}} \end{bmatrix} = \begin{bmatrix} \mathbf{X_{6B}^{m}} & 0 \\ 0 & \mathbf{X_{6B}^{f}} \end{bmatrix} \begin{bmatrix} \mathbf{b_{6B}^{m}} \\ \mathbf{b_{6B}^{f}} \end{bmatrix} + \begin{bmatrix} \mathbf{Z_{6B}^{m}} & 0 \\ 0 & \mathbf{Z_{6B}^{f}} \end{bmatrix} \mathbf{a_{6B}} + \begin{bmatrix} \mathbf{W_{6B}^{m}} & 0 \\ 0 & \mathbf{W_{6B}^{f}} \end{bmatrix} \mathbf{c_{6B}} + \begin{bmatrix} \mathbf{e_{6B}^{m}} \\ \mathbf{e_{6B}^{f}} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{y_{5C}^{m0}} \\ \mathbf{y_{5C}^{f0}} \end{bmatrix} = \begin{bmatrix} \mathbf{X_{5C}^{m}} & 0 \\ 0 & \mathbf{X_{5C}^{f}} \end{bmatrix} \begin{bmatrix} \mathbf{b_{5C}^{m}} \\ \mathbf{b_{5C}^{f}} \end{bmatrix} + \begin{bmatrix} \mathbf{Z_{5C}^{m}} & 0 \\ 0 & \mathbf{Z_{5C}^{f}} \end{bmatrix} \mathbf{a_{5C}} + \begin{bmatrix} \mathbf{W_{5C}^{m}} & 0 \\ 0 & \mathbf{W_{5C}^{f}} \end{bmatrix} \mathbf{c_{5C}} + \begin{bmatrix} \mathbf{e_{5C}^{m}} \\ \mathbf{e_{5C}^{f}} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{y_{6C}^{m0}} \\ \mathbf{y_{6C}^{f0}} \end{bmatrix} = \begin{bmatrix} \mathbf{X_{6C}^{m}} & 0 \\ 0 & \mathbf{X_{6C}^{f}} \end{bmatrix} \begin{bmatrix} \mathbf{b_{6C}^{m}} \\ \mathbf{b_{6C}^{f}} \end{bmatrix} + \begin{bmatrix} \mathbf{Z_{6C}^{m}} & 0 \\ 0 & \mathbf{Z_{6C}^{f}} \end{bmatrix} \mathbf{a_{6C}} + \begin{bmatrix} \mathbf{W_{6C}^{m}} & 0 \\ 0 & \mathbf{W_{6C}^{f}} \end{bmatrix} \mathbf{c_{6C}} + \begin{bmatrix} \mathbf{e_{6C}^{m}} \\ \mathbf{e_{6C}^{f}} \end{bmatrix} \quad (2.2)$$

where $\mathbf{y_{5B}^{m0}}$, $\mathbf{y_{6B}^{m0}}$, $\mathbf{y_{5C}^{m0}}$ and $\mathbf{y_{6C}^{m0}}$ are the vectors of standardized phenotypic records of BW5.B, BW6.B, BW5.C, and BW6.C, respectively, for males; $\mathbf{y_{5B}^{f0}}$, $\mathbf{y_{6B}^{f0}}$, $\mathbf{y_{5C}^{f0}}$ and $\mathbf{y_{6C}^{f0}}$ are the vectors of standardized phenotypic records of BW5.B, BW6.B, BW5.C, and BW6.C, respectively, for females; $\mathbf{b_{5B}^{m}}$, $\mathbf{b_{5B}^{f}}$, $\mathbf{b_{6B}^{m}}$, $\mathbf{b_{6B}^{f}}$, $\mathbf{b_{5C}^{m}}$, $\mathbf{b_{5C}^{f}}$, $\mathbf{b_{6C}^{m}}$ and $\mathbf{b_{6C}^{f}}$ are the vectors of fixed factors as in model (2.1) above; **X**, **Z** and **W** are incidence matrices. The vectors of $\mathbf{a_{5B}}$, $\mathbf{a_{6B}}$, $\mathbf{a_{5C}}$ and $\mathbf{a_{6C}}$ are direct additive genetic effects that were reduced ranks for male and female traits: $\mathbf{a_{5B}}$, $\mathbf{a_{6B}}$, $\mathbf{a_{5C}}$ and $\mathbf{a_{6C}} \sim \mathbf{MVN}[\mathbf{0}, \mathbf{A} \otimes \mathbf{V_a^0}]$, where $\mathbf{V_a^0}$ is the 4x4 covariance matrix; and **A** is the pedigree relationship matrix. The vectors of $\mathbf{c_{5B}}$, $\mathbf{c_{6B}}$, $\mathbf{c_{5C}}$ and $\mathbf{c_{6C}}$ are permanent environmental maternal effects that were reduced ranks for male and female traits: $\mathbf{c_{5B}}$, $\mathbf{c_{6B}}$, $\mathbf{c_{5C}}$ and $\mathbf{c_{6C}} \sim \mathbf{MVN}[\mathbf{0}, \mathbf{I_d} \otimes \mathbf{V_c^0}]$, where $\mathbf{V_c^0}$ is the 4x4 covariance matrix; and $\mathbf{I_d}$ is the identity matrix for dams. The vectors of $\mathbf{e_{5B}^{m}}$, $\mathbf{e_{5B}^{f}}$, $\mathbf{e_{6B}^{m}}$, $\mathbf{e_{6B}^{f}}$, $\mathbf{e_{5C}^{m}}$, $\mathbf{e_{5C}^{f}}$, $\mathbf{e_{6C}^{m}}$ and $\mathbf{e_{6C}^{f}}$ are residuals $\sim \mathbf{MVN}[\mathbf{0}, \mathbf{I} \otimes \mathbf{V_e^0}]$, where $\mathbf{V_e^0}$ is the 8x8 covariance matrix because heterogeneous variances between sexes was applied to residual in the model (2.2); **I** is the identity matrix for individual birds. In matrix $\mathbf{V_e^0}$, covariances between male and female traits, and between B and C traits were zero. Covariance matrices $\mathbf{V_a^0}$, $\mathbf{V_c^0}$ and $\mathbf{V_e^0}$ are at standardized scale.

Variance components were estimated using the PBLUP multivariate model (2.2) for two reasons. One reason is selective genotyping applied to birds in B that might lead to overestimation and bias of variance components estimates with a ssGBLUP model (Cesarani *et al.*, 2019). Another reason is that the complexity of the models and large amount of genotyped animals and non-genotyped animals in the pedigree would make ssGBLUP computationally demanding. The estimation of variance components and EBV prediction with the PBLUP models was carried out using the REML module in the DMUAI procedure of the DMU package (Madsen and Jensen, 2013). Criteria for the convergence of the model were set for the Frobenius norm of the update vector being less than 10^{-5} (Madsen and Jensen, 2013). In addition, estimation of variance components with the PBLUP models were re-run several times using different starting values to check for having reached the global maximum likelihood of the model. The model was converged, and no change in estimates was observed for different starting values.

The standardized variance components of the PBLUP model estimated from the REML module were used to calculate EBV for ssGBLUP multivariate models using the DMU5 procedure of the DMU package (Madsen and Jensen, 2013). Because the use of genomic information led to a relatively dense relationship matrix, calculation of the sparse inverse of LHS was not possible. Therefore, the DMU5 procedure was used to iteratively solve the mixed model equations with the preconditioned conjugate gradient method (Madsen and Jensen, 2013). This procedure, however, does not provide prediction error variance of breeding values. The ssGBLUP models were identical to the PBLUP model (2.2), except that the pedigree relationship matrix **A** was replaced by a combined relationship matrix **H**. The matrix **H** was constructed from the pedigree relationship matrix **A** and genomic relationship matrix **G** with weight value ω = 0.01 (Aguilar *et al.*, 2011; Christensen and Lund, 2010) on the pedigree relationships. The genomic relationship matrix **G** was constructed based on SNP marker data, using method 1 from VanRaden (2008).

Cross-validation

Cross-validation was carried out to evaluate accuracy, bias and dispersion of EBV for genotyped and non-genotyped validation birds in B environment. Validation was based on EBV estimated from full and reduced datasets. The full dataset contained all phenotypic records from TS 1-16 while the reduced dataset contained only phenotypic records from TS 1-12. The reduced dataset was a subset of the full dataset, in which records of 14,187 birds in B and 5,988 birds in C at TS 13-16 were removed. The validation birds were individuals in B environment at TS 13-16. This design was to avoid having two or more generations of validation birds. The validation individuals might be genotyped birds or non-genotyped birds. Cross-validation measures were statistics of EBV of validation birds estimated from the full and reduced datasets (Legarra and Reverter, 2018): correlation ($\rho_{f,r}$) between $\mathbf{EBV_f}$ and $\mathbf{EBV_r}$, difference ($d_{f,r}$)

in means of $\mathbf{EBV_f}$ and $\mathbf{EBV_r}$ and regression slope ($b_{f,r}$) of $\mathbf{EBV_f}$ on $\mathbf{EBV_r}$, where $\mathbf{EBV_f}$ and $\mathbf{EBV_r}$ were vectors of EBV of validation birds estimated from the full and reduced datasets, respectively. Statistics of $\rho_{f,r}$, $d_{f,r}$ and $b_{f,r}$ were indicators of population accuracy, bias and dispersion of EBV, respectively (Legarra and Reverter, 2018). Expectation of $\rho_{f,r}$ is $\frac{acc_r}{acc_f}$ (Legarra and Reverter, 2018), where acc_r is the population accuracy of EBV defined as the correlation between the true breeding values and $\mathbf{EBV_r}$; acc_f is the population accuracy of EBV defined as the correlation between the true breeding values and $\mathbf{EBV_f}$. Expectations of $d_{f,r}$ and $b_{f,r}$ are 0 and 1, respectively (Legarra and Reverter, 2018).

The two validation models were PBLUP and ssGBLUP models. Standardized covariance components that were estimated from the PBLUP model using the full dataset were used to predict EBV in PBLUP and ssGBLUP models using the full dataset or reduced dataset. Vectors of $\mathbf{EBV_f}$ and $\mathbf{EBV_r}$ were at the standardized scale. Scaling does not influence $\rho_{f,r}$ and $b_{f,r}$. However, scaling applied differently to records of each sex means $d_{f,r}$ would have two values for males and females at the normal scale. For comparison between models, $d_{f,r}$ was computed at standardized scale. Standard errors of the statistics $\rho_{f,r}$, $d_{f,r}$ and $b_{f,r}$ were calculated using formula in the appendix.

Rescaling of parameters

Parameters estimated from the PBLUP model (2.2) were on the standardized scale for all BW traits. The estimates were re-scaled back to the normal scale for male and female BW traits. Rescaling of (co)variance matrices and the asymptotic covariance matrices used formula (2.3-2.6):

$$\mathbf{V}_a = \mathbf{T_2}\left(\mathbf{T_1}\,\mathbf{V_a^0}\,\mathbf{T_1'}\right)\mathbf{T_2'} \quad (2.3)$$

$$\mathbf{V}_c = \mathbf{T_2}\left(\mathbf{T_1}\,\mathbf{V_c^0}\,\mathbf{T_1'}\right)\mathbf{T_2'} \quad (2.4)$$

$$\mathbf{V}_e = \mathbf{T_2}\,\mathbf{V_e^0}\,\mathbf{T_2'} \quad (2.5)$$

$$\mathbf{V_I} = \mathbf{T_{V_{I2}}}\left(\mathbf{T_{V_{I1}}}\;\mathbf{V_I^0}\;\mathbf{T_{V_{I1}}'}\right)\mathbf{T_{V_{I2}}'} \quad (2.6)$$

where matrices of direct additive genetic, permanent environmental maternal, residual and asymptotic covariances were $\mathbf{V_a}$, $\mathbf{V_c}$, $\mathbf{V_e}$ and $\mathbf{V_I}$, respectively, at normal scale, and $\mathbf{V_a^0}$, $\mathbf{V_c^0}$, $\mathbf{V_e^0}$ and $\mathbf{V_I^0}$, respectively, at standardized scale. The asymptotic covariance matrix $\mathbf{V_I^0}$ is the inverse of the average of observed and expected information in the REML likelihood (Jensen, 1997) from DMUAI procedures (Madsen and Jensen, 2013). Transforming matrices for formula (2.3-2.6) were:

$$\mathbf{T_1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{T_{V_{I1}}} = \begin{bmatrix} \mathbf{T_1} \otimes \mathbf{T_1} & 0 & 0 \\ 0 & \mathbf{T_1} \otimes \mathbf{T_1} & 0 \\ 0 & 0 & \mathbf{I_e} \otimes \mathbf{I_e} \end{bmatrix},$$

$$\mathbf{T_{V_{I2}}} = \begin{bmatrix} \mathbf{T_2} \otimes \mathbf{T_2} & 0 & 0 \\ 0 & \mathbf{T_2} \otimes \mathbf{T_2} & 0 \\ 0 & 0 & \mathbf{T_2} \otimes \mathbf{T_2} \end{bmatrix}$$

where matrix $\mathbf{I_e}$ is an 8x8 identity matrix. The matrix $\mathbf{I_e} \otimes \mathbf{I_e}$ was adjusted to accounted for non-existent covariances in $\mathbf{V_e^0}$. Matrix $\mathbf{T_2}$ is an 8x8 matrix, of which off-diagonal elements are zero, the diagonal is vector of phenotypic standard deviations with trait orders: male BW5.B, female BW5.B, male BW6.B, female BW6.B, male BW5.C, female BW5.C, male BW6.C and female BW6.C. The phenotypic standard deviations of diagonals from matrices $\mathbf{T_2}$ were computed from univariate model (2.1) for corresponding traits. The asymptotic covariance matrix $\mathbf{V_I}$ was used to compute approximate standard errors for (co)variance component estimates using Taylor series approximation to the asymptotic normal distribution of model parameters.

2.3. Results

Number of records, means, and standard deviations of BW records for broiler chicken raised in B and C environments are shown in Table 2.1. Mean of BW in B was higher than in C for the same week and sex. However, the standard deviation of BW in B was lower than that of BW in C, and thus the coefficient of variation of BW in B was lower. For example, coefficients of variation were 0.098 and 0.174 for male BW5 measured in B and C environments, respectively.

Table 2.1. Descriptive statistics for body weight (BW) records of broiler chicken at 5 and 6 weeks of age for each sex raised in breeding (B) and commercial production (C) environments. Unit of BW was measured in gram.

BW at week	Sex	B environment			C environment		
		Number of records	Mean	Standard deviation	Number of records	Mean	Standard deviation
5	Male	10117	2183	213	7455	1735	302
5	Female	10801	1882	180	7922	1550	248
6	Male	18651	2758	269	3975	2231	364
6	Female	22020	2329	217	4217	1940	290

Standard deviation of BW records increased from 5 to 6 weeks of age. Mean of BW increased by 500-600 grams between week 5 and 6. As a result, the change in coefficients of variation was relatively small. For example, the coefficient of variation for BW records of male birds in B were 0.098 in both weeks 5 and 6. The coefficients of variation for BW records of male broilers in C were 0.174 and 0.163 at week 5 and 6, respectively.

Mean BW in males was higher than in females at corresponding age. The standard deviation in males was also higher than in females, and thus the difference in coefficients of variation between sexes was relatively small. The relative difference in means between males and females was higher for records in B than records in C. The relative difference in standard deviations between males and females was lower for records in B than records in C.

Table 2.1 does not show separated statistics for genotyped and non-genotyped birds in B. However, it was found that genotyped birds in B had higher means and lower standard deviations of BW records than the non-genotyped birds at corresponding weeks of age. For example, mean and standard deviation of BW6 of males were 2814 and 248 for genotyped birds, respectively, and 2707 and 278 for non-genotyped birds, respectively.

Table 2.2. Estimates of direct additive genetic variance (σ_a^2), heritability (h^2) and permanent environmental maternal effect (c^2) estimated from male and female body weight (BW) at week 5 and 6 in breeding (B) and commercial production (C) environments.

BW at week	Sex	B environment			C environment		
		σ_a^2	h^2	c^2	σ_a^2	h^2	c^2
5	Male	10454	0.274	0.033	24984	0.358	0.037
5	Female	7614	0.278	0.033	17469	0.366	0.038
6	Male	17301	0.301	0.034	39544	0.312	0.028
6	Female	11651	0.298	0.034	23831	0.305	0.027
Standard errors in range			0.022–0.024	0.007–0.008		0.033–0.037	0.011–0.013

Additive genetic variances, heritability and permanent environmental maternal effects of B and C traits for male and female BW5-6 were estimated from PBLUP model (2.2) and shown in Table 2.2. Genetic variances of C traits were considerably higher than that of corresponding B traits. The relative differences in variances between B and C traits for male BW5, female BW5, male BW6 and female BW6 were 2.39, 2.29, 2.29 and 2.04, respectively.

Estimates of heritability of C traits tended to be higher than of corresponding B traits. For example, heritability of BW5-6 was 0.27-0.30 and 0.31-0.37 for the same traits measured in B and C environments, respectively. Increasing age from 5 to 6 weeks reduced heritability of C but heritability of B traits tended to increase. Ratio of the permanent environmental maternal variance to the total phenotypic variance for C reduced as week of age increased. Meanwhile, the ratio for B traits tended to increase from week 5 to week 6.

Genetic variances was considerably higher for male BW than female BW when traits in the same environment and at the same week of age were considered. The absolute and relative differences in variances between male and female traits were higher for C traits than B traits. However, the differences in heritability and permanent environmental maternal effect were mostly negligible between male and female traits measured in the same environment and week of age.

Table 2.3. Genetic correlations (above diagonal), permanent environmental maternal correlations (below diagonal)–of body weight (BW) of broiler chicken reared in the breeding environment (B) and in the commercial production environment (C) at 5 and 6 weeks of age.

Environment		B		C	
	BW at week	5	6	5	6
B	5	1	0.956	0.535	0.490
	6	0.953	1	0.497	0.479
C	5	0.690	0.589	1	0.989
	6	0.723	0.628	0.999	1

Standard errors: genetic correlations ±0.010-0.064; permanent environmental maternal correlations ±0.027-0.155

Estimates of variances, heritability and permanent environmental maternal effects were different for male and female traits because the PBLUP model used heterogeneous residual variances and different scaling for sexes. However, genetic correlations and permanent environmental maternal correlations between sexes were assumed to be unity for the same traits. Therefore, a single correlation estimate was obtained for the different sexes combined. Table 2.3 shows genetic correlations and permanent environmental maternal correlations between B and C traits of BW at week 5 and 6 estimated from the PBLUP model.

The genetic correlations between BW traits measured in B and C were 0.48-0.54. The genetic correlation of BW between B and C environments had a reducing tendency as week of age increased from 5-6. The genetic correlations were 0.54 between BW5.B and BW5.C and 0.48 between male BW6.B and BW6.C. Genetic variances in each environment were higher for BW6 traits than BW5 traits. The correlation between BW5 and BW6 tended to be higher in C (0.99) than in B (0.96).

The permanent environmental maternal correlations between B and C traits were 0.59-0.72. The permanent environmental maternal correlations between BW at 5 and 6 weeks, and BW in B and C had similar trend to the genetic correlations. For example, the environmental correlations between B and C were lower for BW6 than for BW5. Permanent environmental maternal correlations between BW5 and BW6 in the same environment were very high but still a bit higher in C than in B.

Table 2.4: Indicators of population accuracy ($\rho_{f,r}$), dispersion ($b_{f,r}$) and bias ($d_{f,r}$) of EBV for genotyped and non-genotyped validation birds in B using pedigree-based BLUP (PBLUP) and single step GBLUP (ssGBLUP) models. Traits for validation were body weights (BW) at 5 and 6 weeks of age in the breeding (B) and commercial production (C) environments.

Model	Validated traits	Genotyped validation birds			Non-genotyped validation birds		
		$\rho_{f,r}$	$b_{f,r}$	$d_{f,r}$	$\rho_{f,r}$	$b_{f,r}$	$d_{f,r}$
PBLUP	BW5.B	0.509	0.898	0.106	0.588	0.874	-0.081
	BW6.B	0.457	0.858	0.129	0.570	0.849	-0.083
	BW5.C	0.616	1.055	0.088	0.678	1.026	-0.052
	BW6.C	0.591	1.024	0.084	0.670	1.014	-0.048
Standard errors in range		0.013-0.015	0.023-0.028	0.004-0.006	0.007-0.008	0.011-0.012	0.002-0.003
ssGBLUP	BW5.B	0.767	0.913	0.026	0.625	0.841	-0.095
	BW6.B	0.791	0.946	0.043	0.653	0.867	-0.091
	BW5.C	0.811	0.937	-0.001	0.758	0.951	-0.076
	BW6.C	0.813	0.939	0.000	0.770	0.965	-0.069
Standard errors in range		0.010-0.011	0.011-0.013	0.003-0.004	0.006-0.008	0.008-0.010	0.002-0.003

Note: $\rho_{f,r}$ is correlation between $\mathbf{EBV_f}$ and $\mathbf{EBV_r}$; $b_{f,r}$ is regression slope of $\mathbf{EBV_f}$ on $\mathbf{EBV_r}$; and $d_{f,r}$ is mean of $\mathbf{EBV_f}$ - $\mathbf{EBV_r}$; vectors of $\mathbf{EBV_f}$ and $\mathbf{EBV_r}$ are breeding values of validation birds in B estimated from full dataset and reduced dataset, respectively.

Statistics of $\rho_{f,r}$, $b_{f,r}$ and $d_{f,r}$ for the PBLUP and ssGBLUP models was calculated based on EBVs estimated from the full and reduced datasets (Table 2.4). These cross-validation measures were computed separately for genotyped and non-genotyped validation birds for EBVs of all BW traits. The use of ssGBLUP increased $\rho_{f,r}$ for all validation birds, and reduced $d_{f,r}$ for genotyped birds compared to PBLUP.

With the PBLUP model, $\rho_{f,r}$ of genotyped birds was lower and $d_{f,r}$ was higher than non-genotyped birds. For BW traits in B, $b_{f,r}$ of genotyped birds was closer to 1 than that of non-genotyped birds. In contrast, for BW traits in C, $b_{f,r}$ of non-genotyped birds was closer to 1. With the PBLUP model, EBV of BW traits in B were deflated, but EBV of BW traits in C were inflated.

With the ssGBLUP model, $\rho_{f,r}$ of genotyped birds was higher than for non-genotyped birds. When the model was changed from PBLUP to ssGBLUP, $\rho_{f,r}$ increased for both genotyped birds and non-genotyped birds. However, the increase in $\rho_{f,r}$ was much larger for genotyped birds than non-genotyped birds. For example, the relative increase in $\rho_{f,r}$ was from 31.7-73.1% for genotyped birds and from 6.3-14.9% for non-genotyped birds. Statistic $d_{f,r}$ of genotyped birds was lower with ssGBLUP than PBLUP model whereas $d_{f,r}$ of non-genotyped birds was higher with ssGBLUP. When the model was changed from PBLUP to ssGBLUP, $b_{f,r}$ of genotyped birds increased for B traits and decreased for C traits, but $b_{f,r}$ of non-genotyped

birds decreased for B traits and increased for C traits. With ssGBLUP model, EBVs were deflated for all traits.

Regardless of model used, $\rho_{f,r}$ of traits in B was lower and $d_{f,r}$ was higher than in C. For example, with PBLUP, $\rho_{f,r}$ of non-genotyped birds was 0.59 for BW5.B and 0.68 for BW5.C, and $d_{f,r}$ of non-genotyped birds was -0.081 for BW5.B and -0.052 for BW5.C.

2.4. Discussion

Genetic parameters for male and female BW at 5 and 6 week of age raised in bio-secure breeding (B) or commercial production (C) environments were estimated using a PBLUP multivariate model. A multivariate ssGBLUP model was used to predict EBV of BW traits in B and C. Cross-validations were carried out to assess population accuracy, bias and dispersion of EBV predictions of C traits for genotyped and non-genotyped birds in B when the PBLUP and ssGBLUP models were used.

Genotype-by-environment interaction

The difference between B and C environments is mainly determined by hygienic conditions. The strict bio-secure conditions of B environment are regulated to prevent worldwide spread of diseases to production farms and to avoid the risk of losing the purebred lines [1]. In contrast, the less strict hygienic conditions of C can result in higher incidence of diseases. For example, purebred birds in large commercial breeding companies are typically disease-free from Salmonellae, Mycoplasma, Leucosis, Avian Influenza and Newcastle disease (Hiemstra and ten Napel, 2013). However, these diseases remains chronic problems in many commercial poultry flocks (De Boeck *et al.*, 2015; EFSA, 2017; Shamim *et al.*, 2015). The difference between B and C may be also related to diet and litter management. Commonly wood shavings are used that are loose, dry, free-flowing in B environment (Hiemstra and ten Napel, 2013). The better litter management lead to substantially lower incidence of food-pad dermatitis in birds of B environment compared to C environment (Kapell *et al.*, 2012). Several indications of GxE due to the difference between B and C environments can be seen in our study such as changed average of performance, re-rankings, heterogeneous variances and different heritability of traits recorded in B and C environments. At the same age, the mean BW in B was higher than the mean BW in C, but the standard deviation of BW in B was lower than for BW in C. This finding is in agreement with Kapell *et al.* (2012) where higher means and lower standard deviations were found for BW in B than in C for four different purebred lines of broiler chicken raised in B and C environments. In N'Dri *et al.* (2007), slow growing broiler chicken raised in B also had significantly higher BW performance than birds raised in C.

In our study, estimates of genetic correlation between BW measured in B and C environments were 0.479-0.535. Statistically, re-ranking refers to genetic correlation different from 1, but in practice, re-ranking is commonly considered important in breeding programs when the correlation is less than 0.8 (Robertson, 1959). The genetic correlation in our study refers to a significant re-ranking of birds in B and C environments. The significant re-ranking was also found in Kapell *et al.* (2012) with the genetic correlations between B and C of 0.46, 0.54, 0.56 and 0.69 for their four studied lines. In N'Dri *et al.* (2007), the genetic correlations were 0.74-0.76. In Lwelamira (2012), the genetic correlation between BW traits of indigenous chicken measured in breeding station and village environments was 0.75-0.76. However, the estimates in Lwelamira (2012) and N'Dri *et al.* (2007) might be slightly overestimated because permanent environmental maternal effect was not included in the model and the datasets had small number of records in only one generation.

With the more challenging environment of C, the higher variances of C traits than B traits were expected, but in our study, the heritability of C traits was also higher than that of the corresponding B traits. Kapell *et al.* (2012) found that in three studied lines, the heritability of C trait (0.32-0.34) was lower than B trait (0.36-0.40) for BW5, while in the remaining line, the heritability of the trait in C (0.36) was higher than the trait in the B environment (0.32). N'Dri *et al.* (2007) found heritability of C traits was 0.54-0.56, and heritability of B traits was 0.56. In addition, as week of age increased from 5 to 6, heritability of B traits tended to increase, but heritability of C traits decreased. There is no clear explanation to the opposite trend of the two traits. As age increased from 5 to 6 weeks, the traits measured in B and C environments had lower genetic correlation. To our knowledge, there are no prior studies on poultry having reported a trend in the genetic correlation between B and C environments with increasing age.

The production environment also had effects on the maternal environmental effects. Permanent environmental maternal correlations between BW measured in B and C environments were significantly below unity. Permanent environmental maternal effect of C traits became smaller with increasing age from 5 to 6 weeks. In contrast, permanent environmental maternal effect of B traits tended to increase as age increased from 5 to 6 weeks. With increasing age of poultry, the reduction of the permanent environment maternal effect on BW can be found in several studies (Barbieri *et al.*, 2015; Dana *et al.*, 2011; Jasouri *et al.*, 2017; Maniatis *et al.*, 2013). However, some studies (Begli *et al.*, 2016; Mebratie *et al.*, 2017) also show increasing trends around the age of 5-6 week. Begli *et al.* (2016) showed that the ratio of the permanent environmental maternal variance to the total phenotypic variance for BW increased slightly from 0.10 at week 2 to 0.12 at week 6, and then reduced to 0.07 at week 10. Mebratie *et al.* (2017)

showed that the maternal effect on BW of broiler chicken raised in B had an increasing trend as birds aged from *t-7*, *t-4* to *t* days.

Effects of the production environment on age-by-genotype interactions were not clear, but the genetic correlation between BW5 and BW6 tended to be lower for B traits than C traits. Effects of the production environment on sex-by-genotype interaction related mainly to the relative difference in variances between male and female BW traits because the genetic correlation between male and female BW within each environment was assumed to be 1 in the multivariate model (2.2). This assumption came from the preliminary results that sex-by-genotype interaction only led to the scaling effect between male and female BW.

With a strong GxE interaction found in this study, selection for performance under commercial conditions will greatly increase response to selection for BW in broiler breeding programs when birds are phenotype-tested in both B and C environments (Chu *et al.*, 2018; Mulder and Bijma, 2005). In contrast to the breeding programs that test birds in B environment only, the breeding programs that test birds in both the two environments can exploit the re-ranking of birds in B and C, the larger genetic variances of C traits (than B traits) and the higher heritability of C traits (than B traits). Chu *et al.* (2018) shows that with the genetic correlation between B and C traits of 0.5, the scheme that had 70% and 30% birds placed to B and C environments, respectively, for phenotype testing had substantially larger genetic gains than the scheme that had all birds tested in B environment.

Benefits of genomic information to prediction of breeding values

The increase in population accuracy of EBV for genotyped birds has been shown in number of studies (Aguilar *et al.*, 2010; Aguilar *et al.*, 2011; Andonov *et al.*, 2017; Chen *et al.*, 2011a; Chen *et al.*, 2011b; Christensen and Lund, 2010; Christensen *et al.*, 2012; Legarra *et al.*, 2009; Lourenco *et al.*, 2013; Misztal *et al.*, 2009), in which the population accuracy was reflected by correlation between EBV and corrected phenotypes. Simulation studies also showed that correlations between true breeding values and genomic EBV of genotyped individuals were significantly higher with ssGBLUP than PBLUP models (Andonov *et al.*, 2017; Christensen and Lund, 2010; Lourenco *et al.*, 2013; Putz *et al.*, 2018). Applications of ssGBLUP to improve accuracy of selection have been well-documented in studies on chicken (Alemu *et al.*, 2016; Chen *et al.*, 2011a; Chen *et al.*, 2011b; Momen *et al.*, 2017; Wolc *et al.*, 2011), cattle (Gao *et al.*, 2018; Lee *et al.*, 2017; Li *et al.*, 2016; Lourenco *et al.*, 2015) and pig (Christensen *et al.*, 2012; Guo *et al.*, 2015; Putz *et al.*, 2018; Xiang *et al.*, 2016). However, the extent of benefits from genomic information for the accuracy of prediction has not been reported for breeding programs where sib-testing is used due to GxE interactions. In these breeding programs, the ultimate goal is to increase genetic gains of performance in

C for selection candidates resided in B, thus only accuracy of EBVs of BW traits in C matters. Our discussion focus primarily on accuracy of EBVs of C traits for the selection candidate birds in B that do not have own record of the C traits.

The statistics of $\rho_{f,r}$, $d_{f,r}$ and $b_{f,r}$ were indicators of population accuracy, bias and dispersion of EBV, respectively (Legarra and Reverter, 2018). Statistic $\rho_{f,r}$ is in ratio of accuracies. Prediction of EBV is expected to be more accurate when $\rho_{f,r}$ was higher. Statistic $\rho_{f,r}$ is expected to be always lower than 1. Prediction of EBV is more biased when the value of $d_{f,r}$ deviates further from 0. The EBV would be more inflated or deflated when the value of $b_{f,r}$ deviates further from 1. The use of combined pedigree and genomic information in ssGBLUP increased substantially the population accuracy of EBV for genotyped validation birds compared to the use of only pedigree in PBLUP. Genomic information explains better relationships between individuals compared to pedigree information. Thus genomic information can be beneficial for an efficient "flow" of information from C to validation birds in B in several ways. With pedigree information, only up to 50% of the total genetic variance of the C traits is exploited for the prediction of EBV of the candidates in a GxE sib-test. With genomic information, the percentage can be higher because the realized genomic relationships between full-sib individuals can range from 0.27 to 0.70 for broilers (Hawken *et al.*, 2015). The prediction of PBLUP for validation birds typically treats phenotypic performances of their full-sibs in C as an average information whereas phenotypic performances of the full-sibs in C are treated individually in genomic prediction of EBV. In addition, information from related, more distantly related and even unrelated animals can be exploited in genomic prediction when genomic markers are in linkage disequilibrium with genotypes at casual loci (Daetwyler *et al.*, 2013). The Mendelian sampling terms are exploited better with genomic information than pedigree information (Daetwyler *et al.*, 2013; Daetwyler *et al.*, 2007; Hayes *et al.*, 2009), and thus accuracy of prediction from genomic information increases compared to the accuracy from pedigree information.

Most studies (Aguilar *et al.*, 2010; Aguilar *et al.*, 2011; Andonov *et al.*, 2017; Chen *et al.*, 2011a; Chen *et al.*, 2011b; Christensen and Lund, 2010; Christensen *et al.*, 2012; Legarra *et al.*, 2009; Lourenco *et al.*, 2013; Misztal *et al.*, 2009) reported regression slope of corrected phenotypes on EBV as bias, and they showed an improvement in bias with genotyped animals. However, methodology of cross-validation from those studies is different from our study on e.g. validation animals and definition of bias. In our study, when the model was changed from PBLUP to ssGBLUP, dispersion or regression slope of $\mathbf{EBV_f}$ on $\mathbf{EBV_r}$ for genotyped birds was improved with B traits, but not C traits, and bias or difference in mean between $\mathbf{EBV_f}$ on $\mathbf{EBV_r}$ for genotyped birds was improved with both B and C traits. In literature, when the model was changed from PBLUP to ssGBLUP, an additional accuracy of EBV was also reported for non-genotyped

individuals (Christensen *et al.*, 2012; Gao *et al.*, 2018; Guo *et al.*, 2015; Xiang *et al.*, 2016), but bias of prediction of these birds increased (Gao *et al.*, 2018; Guo *et al.*, 2015; Xiang *et al.*, 2016). The cross-validation in these studies was based on corrected phenotypes and EBV. Nonetheless, the results are in agreement with our study, showing that accuracy of EBV of non-genotyped birds increased, but bias of the prediction also increased.

Validation groups of genotyped and non-genotyped birds had different population accuracy of EBV even with the PBLUP model, which might be related to non-random division of validation birds into the groups. In our dataset, mean of BW records in B was higher for genotyped birds than non-genotyped ones, but the standard deviation of B records was lower for genotyped birds than non-genotyped ones. All parents had genotyping information. This indicated that a selective genotyping strategy was applied for birds in B. Within validation birds, the genotyped ones that are from the top of the distribution have higher relationships between individuals than the non-genotyped birds. From the numerator relationship matrix that was calculated from all animals in the pedigree, the average additive genetic relationship was 0.024 between validation genotyped birds and 0.019 between validation non-genotyped birds. The higher average relationships between genotyped animals lead to less variation in EBV estimated from PBLUP, and thus lower population accuracy for genotyped birds.

Methodology

When genomic information was utilized and evaluated in a GxE sib-testing breeding program for broilers, several challenges were faced in our study. Estimation of variance components was challenging with a high number of genotyped birds, computation-demanding multi-trait models, and selective genotyping. Cross-validation strategies were also not clear for the situation where selection candidates reside in B and do not have C performance. Cross-validation is commonly based on correlation between phenotypes that are corrected for fixed effects from PBLUP model using the full dataset and EBV that are estimated from PBLUP or ssGBLUP model using the reduced dataset (Christensen *et al.*, 2012). However, corrected phenotypes for C traits were not available for validation birds in B. To deal with those challenges, our study used PBLUP model for estimation of variance components. Statistics of EBV that were estimated from full and reduced data for validation birds in B were used for validation measures e.g. indicators of population accuracy, bias and dispersion of EBV following the derivation of Legarra and Reverter (2018). Strictly, the statistic $\rho_{f,r}$ is not population accuracy, but a direct indicator of population accuracy in ratio of accuracies (Legarra and Reverter, 2018). The statistic $\rho_{f,r}$ describes the increase in population accuracy of EBV from reduced data to full data. We observed higher values of $\rho_{f,r}$ for B traits than C traits, but these values are not comparable for different traits. Without realizing the ratio form of statistic $\rho_{f,r}$, Putz *et al.*

(2018) reported a poor performance of this statistic $\rho_{f,r}$ in estimating accuracy of EBV. To estimate population accuracy from the statistic $\rho_{f,r}$, we need prediction error variances and covariances, and genetic variance at equilibrium in a population under selection (Legarra and Reverter, 2018). However, instead of transforming to population accuracy, the statistic $\rho_{f,r}$ can be used directly for comparisons of accuracy between competing statistical models or between subsets of animals in a population. Analytical properties of $\rho_{f,r}$, $d_{f,r}$ and $b_{f,r}$ were presented in Legarra and Reverter (2018). Using data of a simulated population and empirical data of a Brahman beef cattle population, the authors showed very good agreement between the common cross-validation that is based on corrected phenotypes and EBV estimated from reduced dataset and the new cross-validation that is based on EBV estimated from reduced and full datasets.

The statistic $\rho_{f,r}$ is an indicator of population accuracy, not an indicator of individual accuracy (Legarra and Reverter, 2018). Individual accuracy or model-based accuracy that could be obtained from prediction error variances was not used for model comparisons in our study, because obtaining prediction error variances was not possible for the ssGBLUP model. In addition, individual accuracy reflects "the credibility of an individual EBV" or "a measure of the standard error of prediction of an individual EBV" (Bijma, 2012). Population accuracy reflects "the correlation between true breeding values and EBV among the candidates for selection, which is a property of a population, not of an individual" (Bijma, 2012). For example, when only parent average is known, individual accuracy of full-sibs is up to 0.71. However, the predicted differences between full sibs have zero accuracy, or population accuracy among full-sibs is zero because full-sibs have the same parent average. Individual accuracy should be used for individual decisions, but population accuracy should be used for the choice of model and the assessment of genetic gain (Legarra and Reverter, 2018).

Sex-by-genotype interaction has been investigated for different traits in broiler chicken (Mebratie *et al.*, 2017; Mignon-Grasteau *et al.*, 2000; van der Heide *et al.*, 2016), other poultry (Chapuis *et al.*, 1996; Mignon-Grasteau *et al.*, 1998), pig and cattle (Retallick *et al.*, 2015; van der Heide *et al.*, 2016). These studies showed that variances between sexes differed by a factor of 2 or more, but very little re-ranking between sexes was found with genetic correlations between male and female traits above 0.85. With high genetic correlation, the implementation of bivariate model treating sexes as different traits could encounter convergence problems in variance component estimation. However, when the existence of heterogeneous variance for sexes were not accounted for in the model, a serious re-ranking could occur and thus lead to reduced response to selection (Cardoso *et al.*, 2007). Failure to account for different variances between sexes could also lead to bias in variance components and predicted breeding values

(Thompson, 2008). Therefore, a relatively simple standardization was applied to male and female records to model heterogeneous variances between sexes without affecting convergence of the model. This standardization also has been applied in Chu *et al.* (2019).

Data on BW5-6 of birds in C for our study is the same data as for Chu *et al.* (2019), but the models used were different. Permanent environmental maternal effects in our study were higher than those reported in Chu *et al.* (2019). The six-trait model in Chu *et al.* (2019) used BW1-2 and weekly weight gains from week 2-6 to model BW1-6, and this linear transformation led to the same inferences, but have much better convergence properties. However, because of the trade-off between the complexity of the six-trait model and convergence challenge, permanent environmental maternal effects were not included into the model for weekly weight gain from 5 to 6. Therefore, Chu *et al.* (2019) might overestimate direct additive genetic effects and underestimate the maternal effects. The model used in our study was less complex with four-trait model. In addition, to facilitate convergence, our study had slightly lower convergence criteria from the recommended value of 10^{-6} (Madsen and Jensen, 2013) to 10^{-5} for the Frobenius norms of the update vector and the gradient vector. Therefore, to ensure the global maximum likelihood of the model were reached, estimation of variance components with the PBLUP models were re-run several times with different starting values.

2.5. Conclusions

Genetic parameters were estimated for male and female body weight (BW) at 5 and 6 week of age raised in breeding bio-secure (B) and production commercial environments (C). Several indications of interaction between genotype and testing environment (B and C) were found including different average performance, correlations significantly lower than 1, heterogeneous variances and different heritability for B and C traits. In addition, genomic information on birds in both B and C was used for prediction of EBV of birds in B for BW traits measured in B and C environments. To evaluate the prediction, cross-validation statistics of EBV were estimated, namely population accuracy, bias and dispersion of EBV from reduced and full datasets. It was found that the use of combined pedigree and genomic information in ssGBLUP substantially increased population accuracy of EBV for genotyped birds compared to the use of only pedigree in PBLUP. The increase in accuracy of EBV also occurred for non-genotyped birds, but bias of EBV prediction increased for non-genotyped birds. In summary, the difference between B and C environments leads to a strong GxE interaction with genetic correlation of 0.479-0.535 between BW traits of broilers measured in B and C environments. In order to ensure maximum genetic gain under commercial conditions, breeding programs should establish recording systems under commercial

circumstances to provide their customers with genotypes well adapted to the commercial environment. Compared to the use of pedigree only, the use of combined pedigree and genomic information increases substantially population accuracy of EBV for genotyped birds that resides in B.

Appendix:

Standard errors of the statistics $\rho_{f,r}$, $d_{f,r}$ and $b_{f,r}$ were calculated. Linear model for vector **EBV$_f$** (called $\hat{u}_{f_i}$ in scalar notation) on explanatory variable of vector **EBV$_r$** (called $\hat{u}_{r_i}$ in scalar notation) was assumed: $\hat{u}_{f_i} = \alpha + b_{f,r}\hat{u}_{r_i} + \varepsilon_i$, where *i =1, 2, … n*; *n* is the length of vector **EBV$_f$** or **EBV$_r$**; α is the constant intercept; $b_{f,r}$ is the regression slope; and ε_i is the residual term.

The standard error of the estimated regression slope ($\widehat{b_{f,r}}$) is:

$$SE(\widehat{b_{f,r}}) = \sqrt{\frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sum_{i=1}^{n}(\hat{u}_{r_i} - \overline{\hat{u}_r})^2}}$$

The standard error of the estimated correlation ($\widehat{\rho_{f,r}}$) between $\hat{u}_{f_i}$ and $\hat{u}_{r_i}$ is:

$$SE(\widehat{\rho_{f,r}}) = \sqrt{\frac{1 - \widehat{\rho_{f,r}}^2}{n-2}}$$

The standard error of the difference ($\widehat{d_{f,r}}$) between $\hat{u}_{f_i}$ and $\hat{u}_{r_i}$:

$$SE(\widehat{d_{f,r}}) = \frac{\sqrt{\sum_{i=1}^{n}(\hat{u}_{f_i} - \hat{u}_{r_i} - (\overline{\hat{u}_f} - \overline{\hat{u}_r}))^2}}{n}$$

References:

Aguilar, I, Misztal, I, Johnson, DL, Legarra, A, Tsuruta, S & Lawlor, TJ 2010, 'Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score', *Journal of Dairy Science,* vol. 93, no. 2, pp. 743-752.

Aguilar, I, Misztal, I, Legarra, A & Tsuruta, S 2011, 'Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation', *Journal of Animal Breeding and Genetics,* vol. 128, no. 6, pp. 422-428.

Alemu, SW, Calus, MPL, Muir, WM, Peeters, K, Vereijken, A & Bijma, P 2016, 'Genomic prediction of survival time in a population of brown laying hens showing cannibalistic behavior', *Genetics Selection Evolution,* vol. 48, no. 1, pp. 68.

Andonov, S, Lourenco, DAL, Fragomeni, BO, Masuda, Y, Pocrnic, I, Tsuruta, S & Misztal, I 2017, 'Accuracy of breeding values in small genotyped populations using different sources of external information-A simulation study', *J Dairy Sci,* vol. 100, no. 1, pp. 395-401.

Barbieri, A, Ono, RK, Cursino, LL, Farah, MM, Pires, MP, Bertipaglia, TS, Pires, AV, Cavani, L, Carreno, LO & Fonseca, R 2015, 'Genetic parameters for body weight in meat quail', *Poult Sci,* vol. 94, no. 2, pp. 169-71.

Begli, HE, Torshizi, RV, Masoudi, AA, Ehsani, A & Jensen, J 2016, 'Longitudinal analysis of body weight, feed intake and residual feed intake in F_2 chickens', *Livestock Science,* vol. 184, no. 28-34.

Bijma, P 2012, 'Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations', *J Anim Breed Genet,* vol. 129, no. 5, pp. 345-58.

Boligon, AA, Long, N, Albuquerque, LG, Weigel, KA, Gianola, D & Rosa, GJM 2012, 'Comparison of selective genotyping strategies for prediction of breeding values in a population undergoing selection', *Journal of Animal Science,* vol. 90, no. 13, pp. 4716-4722.

Cardoso, FF, Rosa, GJ & Tempelman, RJ 2007, 'Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of postweaning gain of Nelore-Hereford cattle', *J Anim Sci,* vol. 85, no. 4, pp. 909-18.

Cesarani, A, Pocrnic, I, Macciotta, NPP, Fragomeni, BO, Misztal, I & Lourenco, DAL 2019, 'Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies', *Journal of Animal Breeding and Genetics,* vol. 136, no. 1, pp. 40-50.

Chapuis, H, Tixier-Boichard, M, Delabrosse, Y & Ducrocq, V 1996, 'Multivariate restricted maximum likelihood estimation of genetic parameters for production traits in three selected turkey strains', *Genetics Selection Evolution,* vol. 28, no. 3, pp. 299.

Chen, C, Misztal, I, Aguilar, I, Tsuruta, S, Aggrey, S, Wing, T & Muir, W 2011a, 'Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens', *Journal of animal science,* vol. 89, no. 1, pp. 23-28.

Chen, CY, Misztal, I, Aguilar, I, Legarra, A & Muir, WM 2011b, 'Effect of different genomic relationship matrices on accuracy and scale', *J Anim Sci,* vol. 89, no. 9, pp. 2673-9.

Christensen, OF & Lund, MS 2010, 'Genomic prediction when some animals are not genotyped', *Genetics Selection Evolution,* vol. 42, no. 1, pp. 2.

Christensen, OF, Madsen, P, Nielsen, B, Ostersen, T & Su, G 2012, 'Single-step methods for genomic evaluation in pigs', *Animal,* vol. 6, no. 10, pp. 1565-71.

Chu, TT, Alemu, SW, Norberg, E, Sørensen, AC, Henshall, J, Hawken, R & Jensen, J 2018, 'Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 52.

Chu, TT, Madsen, P, Norberg, E, Wang, L, Marois, D, Henshall, J & Jensen, J 2019, 'Genetic analysis on body weight at different ages in broiler chicken raised in commercial environment'. *In* '*Genotype by environment interaction in poultry breeding programs*' (T. T. Chu, ed.), PhD thesis- Department of of Molecular Biology and Genetics. Aarhus University, Denmark.

Daetwyler, HD, Calus, MP, Pong-Wong, R, de Los Campos, G & Hickey, JM 2013, 'Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking', *Genetics,* vol. 193, no. 2, pp. 347-65.

Daetwyler, HD, Villanueva, B, Bijma, P & Woolliams, JA 2007, 'Inbreeding in genome-wide selection', *J Anim Breed Genet,* vol. 124, no. 6, pp. 369–376.

Dana, N, vander Waaij, EH & van Arendonk, JAM 2011, 'Genetic and phenotypic parameter estimates for body weights and egg production in Horro chicken of Ethiopia', *Tropical Animal Health and Production,* vol. 43, no. 1, pp. 21-28.

De Boeck, C, Kalmar, I, Dumont, A & Vanrompay, D 2015, 'Longitudinal monitoring for respiratory pathogens in broiler chickens reveals co-infection of Chlamydia psittaci and Ornithobacterium rhinotracheale', *J Med Microbiol,* vol. 64, no. Pt 5, pp. 565-74.

EFSA, EFSA 2017, 'The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016', *EFSA Journal,* vol. 15, no. 12, pp. 5077.

Gao, H, Koivula, M, Jensen, J, Strandén, I, Madsen, P, Pitkänen, T, Aamand, GP & Mäntysaari, EA 2018, 'Short communication: Genomic prediction using different single-step methods in the Finnish red dairy cattle population', *Journal of Dairy Science,* vol. 101, no. 11, pp. 10082-10088.

Guo, X, Christensen, OF, Ostersen, T, Wang, Y, Lund, MS & Su, G 2015, 'Improving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method', *J Anim Sci,* vol. 93, no.

Hawken, R, Sapp, R, Okimoto, R, Chen, J, Borg, R, Huang, C-H, Wing, T, Wang, C, Henshall, J & Abrahamsen, M 2015, 'The opportunities and challenges of integrating genomics in a broiler breeding program', *Proc. Assoc. Advmt. Anim. Breed. Genet.,* vol. 21, no. 266-269.

Hayes, BJ, Visscher, PM & Goddard, ME 2009, 'Increased accuracy of artificial selection by using the realized relationship matrix', *Genet Res,* vol. 91, no.

Hiemstra, SJ & ten Napel, J 2013, 'Study of the impact of genetic selection on the welfare of chickens bred and kept for meat production', Final report of a project commissioned by the European Commission (DG SANCO 2011/12254).

Jasouri, M, Zamani, P & Alijani, S 2017, 'Dominance genetic and maternal effects for genetic evaluation of egg production traits in dual-purpose chickens', *Br Poult Sci,* vol. 58, no. 5, pp. 498-505.

Jensen, J 1997, 'Residual maximum likelihood estimation of (co) variance components in multivariate mixed linear models using average information', *Journal of the Indian Society of Agricultural Statistics,* vol. 49, no. 215-236.

Kapell, DN, Hill, WG, Neeteson, AM, McAdam, J, Koerhuis, AN & Avendano, S 2012, 'Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments', *Poultry Science,* vol. 91, no. 3, pp. 565-74.

Lee, J, Cheng, H, Garrick, D, Golden, B, Dekkers, J, Park, K, Lee, D & Fernando, R 2017, 'Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle', *Genetics Selection Evolution,* vol. 49, no. 1, pp. 2.

Legarra, A, Aguilar, I & Misztal, I 2009, 'A relationship matrix including full pedigree and genomic information', *J Dairy Sci,* vol. 92, no.

Legarra, A & Reverter, A 2018, 'Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 53.

Li, X, Lund, MS, Zhang, Q, Costa, CN, Ducrocq, V & Su, G 2016, 'Short communication: Improving accuracy of predicting breeding values in Brazilian Holstein population by adding data from Nordic and French Holstein populations', *Journal of Dairy Science,* vol. 99, no. 6, pp. 4574-4579.

Lourenco, DA, Misztal, I, Wang, H, Aguilar, I, Tsuruta, S & Bertrand, JK 2013, 'Prediction accuracy for a simulated maternally affected trait of beef cattle using different genomic evaluation models', *J Anim Sci,* vol. 91, no. 9, pp. 4090-8.

Lourenco, DA, Tsuruta, S, Fragomeni, BO, Masuda, Y, Aguilar, I, Legarra, A, Bertrand, JK, Amen, TS, Wang, L, Moser, DW & Misztal, I 2015, 'Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus', *J Anim Sci,* vol. 93, no. 6, pp. 2653-62.

Lwelamira, J 2012, 'Genotype-Environment (GXE) Interaction for Body Weights for Kuchi Chicken Ecotype of Tanzania Reared Under Intensive and Extensive Management', *Global Journal of Medical Research,* vol. 12, no. 5, pp.

Madsen, P & Jensen, J 2013, 'DMU: a user's guide'. "*A package for analysing multivariate mixed models, Version 6, release 5.2. Home page: http://dmu.agrsci.dk/*",

Maniatis, G, Demiris, N, Kranis, A, Banos, G & Kominakis, A 2013, 'Model comparison and estimation of genetic parameters for body weight in commercial broilers', *Canadian journal of animal science,* vol. 93, no. 1, pp. 67-77.

Mebratie, W, Shirali, M, Madsen, P, Sapp, R, Hawken, R & Jensen, J 2017, 'The effect of selection and sex on genetic parameters of body weight at different ages in a commercial broiler chicken population', *Livestock Science,* vol. 204, no. 78-87.

Meuwissen, THE, Hayes, BJ & Goddard, ME 2001, 'Prediction of total genetic value using genome-wide dense marker maps', *Genetics,* vol. 157, no. 4, pp. 1819-29.

Mignon-Grasteau, S, Beaumont, C, Poivey, J-P & De Rochambeau, H 1998, 'Estimation of the genetic parameters of sexual dimorphism of body weight in'label'chickens and Muscovy ducks', *Genetics Selection Evolution,* vol. 30, no. 5, pp. 481.

Mignon-Grasteau, S, Piles, M, Varona, L, De Rochambeau, H, Poivey, J, Blasco, A & Beaumont, C 2000, 'Genetic analysis of growth curve parameters for male and female chickens resulting from selection on shape of growth curve', *Journal of Animal Science,* vol. 78, no. 10, pp. 2515-2524.

Misztal, I, Legarra, A & Aguilar, I 2009, 'Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information', *J Dairy Sci,* vol. 92, no.

Momen, M, Mehrgardi, AA, Sheikhy, A, Esmailizadeh, A, Fozi, MA, Kranis, A, Valente, BD, Rosa, GJM & Gianola, D 2017, 'A predictive assessment of genetic correlations between traits in chickens using markers', *Genetics Selection Evolution,* vol. 49, no. 1, pp. 16.

Mulder, H & Bijma, P 2005, 'Effects of genotype× environment interaction on genetic gain in breeding programs', *Journal of Animal Science,* vol. 83, no. 1, pp. 49-61.

N'Dri, AL, Sellier, N, Tixier-Boichard, M, Beaumont, C & Mignon-Grasteau, S 2007, 'Genotype by environment interactions in relation to growth traits in slow growing chickens', *Genetics Selection Evolution,* vol. 39, no. 5, pp. 513-528.

Putz, AM, Tiezzi, F, Maltecca, C, Gray, KA & Knauer, MT 2018, 'A comparison of accuracy validation methods for genomic and pedigree-based predictions of swine litter size traits using Large White and simulated data', *J Anim Breed Genet,* vol. 135, no. 1, pp. 5-13.

Retallick, K, Bormann, J, Weaber, R, MacNeil, M, Bradford, H, Freetly, HC, Snelling, W, Thallman, M, Moser, D & Kuehn, L 2015, 'Genetic variance and covariance components for feed intake, average daily gain, and postweaning gain in growing beef cattle', *Kansas Agricultural Experiment Station Research Reports,* vol. 1, no. 1, pp. 9.

Robertson, A 1959, 'The sampling variance of the genetic correlation coefficient', *Biometrics,* vol. 15, no. 3, pp. 469-485.

Shamim, A, ul Hassan, M, Yousaf, A, Iqbal, MF, Zafar, MA, Siddique, RM & Abubakar, M 2015, 'Occurrence and identification of Emeria species in broiler rearing under traditional system', *Journal of Animal Science and Technology,* vol. 57, no. 41.

Thompson, R 2008, 'Estimation of quantitative genetic parameters', *Proceedings of the Royal Society of London B: Biological Sciences,* vol. 275, no. 1635, pp. 679-686.

van der Heide, E, Lourenco, D, Chen, C-Y, O. Herring, W, L. Sapp, R, Moser, D, Tsuruta, S, Masuda, Y, Ducro, B & Misztal, I 2016, 'Sexual dimorphism in livestock species selected for economically important traits', *Journal of Animal science,* vol. 94, no. 9, pp. 3684-3692.

VanRaden, PM 2008, 'Efficient methods to compute genomic predictions', *J Dairy Sci,* vol. 91, no.

Wolc, A, Stricker, C, Arango, J, Settar, P, Fulton, JE, O'Sullivan, NP, Preisinger, R, Habier, D, Fernando, R, Garrick, DJ, Lamont, SJ & Dekkers, JCM 2011, 'Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model', *Genetics Selection Evolution,* vol. 43, no. 1, pp. 5.

Xiang, T, Nielsen, B, Su, G, Legarra, A & Christensen, OF 2016, 'Application of single-step genomic evaluation for crossbred performance in pig', *Journal of animal science,* vol. 94, no. 3, pp. 936-948.

3

Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken

Thinh T. Chu[1,2], Setegn W. Alemu[1], Elise Norberg[1,3], Anders C. Sørensen[1], John Henshall[4], Rachel Hawken[4], Just Jensen[1]*

[1] Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark

[2] Wageningen University & Research Animal Breeding and Genomics, 6709 PG Wageningen, The Netherlands

[3] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[4] Cobb-Vantress Inc., Siloam Springs, Arkansas 72761-1030, United States of America

Corresponding author's email: chu.thinh@mbg.au.dk

Published in *Genetics Selection Evolution*, 2018, 50:52

https://doi.org/10.1186/s12711-018-0430-x

Abstract

Background: A breeding program for commercial broiler chicken that is carried out under strict biosecure conditions can show reduced genetic gain due to genotype by environment interactions (G×E) between bio-secure (B) and commercial production (C) environments. Accuracy of phenotype-based best linear unbiased prediction (BLUP) of breeding values of selection candidates using sib-testing in C is low. Genomic prediction based on dense genetic markers may improve accuracy of selection. Stochastic simulation was used to explore the benefits of genomic selection in breeding schemes for broiler chicken that include birds in both B and C for assessment of phenotype.

Results: When genetic correlations (r_g) between traits measured in B and C were equal to 0.5 and 0.7, breeding schemes with 15, 30 and 45% of birds assessed in C resulted in higher genetic gain for performance in C compared to those without birds in C. The optimal proportion of birds phenotyped in C for genetic gain was 30%. When the proportion of birds in C was optimal and genotyping effort was limited, allocating 30% of the genotyping effort to birds in C was also the optimal genotyping strategy for genetic gain. When r_g was equal to 0.9, genetic gain for performance in C was not improved with birds in C compared to schemes without birds in C. Increasing the heritability of traits assessed in C increased genetic gain significantly. Rates of inbreeding decreased when the proportion of birds in C increased because of a lower selection intensity among birds retained in B and a reduction in the probability of co-selecting close relatives.

Conclusions: If G×E interactions (r_g of 0.5 and 0.7) are strong, a genomic selection scheme in which 30% of the birds hatched are phenotyped in C has larger genetic gain for performance in C compared to phenotyping all birds in B. Rates of inbreeding decreased as the proportion of birds moved to C increased from 15 to 45%.

Keywords: GxE; breeding program; genomic selection; commercial broiler chicken; stochastic simulation

3.1. Introduction

In commercial broiler chicken breeding companies, purebred lines are kept under strict bio-secure environmental conditions (B) to avoid the risk of losing the lines and to prevent worldwide spread of diseases (Hiemstra and ten Napel, 2013). In contrast, birds in the commercial production environment (C) are living under less strict hygienic conditions where diseases might cause poor performance, death or dysfunction of birds. For example, diseases caused by pathogenic *mycoplasma* are still chronical problems in many commercial poultry flocks (Michiels *et al.*, 2016) whereas these pathogens are completely

eradicated in giant commercial breeding companies like Aviagen and Cobb-Vantress (Hiemstra and ten Napel, 2013). The differences between B and C environments can affect phenotypic expression of traits, which might change genetic ranking of breeding birds in the way that the best individual in B might not be the best in C, i.e. genotype by environment interaction (GxE) is expected in this situation. Genotype by environment interactions due to differences between B and C has been found for a number of traits (Kapell *et al.*, 2012; Long *et al.*, 2008; N'Dri *et al.*, 2007; Ye *et al.*, 2006). Kapell *et al.* (2012) reported substantial GxE for bodyweight and foot-pad dermatitis when phenotypes of purebred broiler chicken were recorded in both B and C. They found that genetic correlations (r_g) between traits measured in B and C were 0.46-0.69 for body weight and 0.78-0.82 for foot-pad dermatitis. N'Dri *et al.* (2007) found that r_g were 0.74-0.82 for body weight and 0.84-0.93 for meat quality traits. Long *et al.* (2008) and Ye *et al.* (2006) also found significant GxE interaction for body weight, mortality and other performance traits measured in B and C environments. Therefore, a breeding program carried out under disease-free condition of B might show reduced genetic gain due to GxE, since only gains obtained in C have substantial economic value.

To improve the performance of commercial animals in the presence of GxE, classical method uses sib-testing for phenotypes in both B and C environments and pedigree-based BLUP for prediction of breeding values. This method has been demonstrated in pig and cattle breeding programs (Bijma and Arendonk, 1998; Jiang and Groen, 1999; Mulder and Bijma, 2005). These studies assumed a fixed number of animals in B, but did not limit the number of animals in C. They found that a higher genetic gain can be achieved with phenotypes from both B and C environments. However, when r_g was 0.9, the extra genetic gain was small and large amounts of information from C were needed in order to significantly increase genetic gains. These studies did not investigate the situations where number of animals available for phenotype testing is limited. In broiler chicken breeding programs, the number of hens mated to a rooster is limited, and practical facilities only allow for a limited number of offspring per hen to be hatched at the same time. In addition, in such program, birds in C cannot be brought back to B environment due to bio-security restrictions, and thus cannot be selection candidates but only be used as sources of information for relatives (sibs) in B. In the classical method of sib testing using pedigree, the prediction of breeding values for selection candidate birds in B has low accuracy due to lack of information on Mendelian sampling terms.

Alternatively, genomic prediction based on dense marker genotypes can be an interesting option due to better modeling of relationships between individuals, and a better prediction of the Mendelian sampling terms (Daetwyler *et al.*, 2007; Hayes *et al.*, 2009). It has been illustrated that accuracy of selection can be

improved considerably by genomic selection using high density markers (Daetwyler *et al.*, 2007; Meuwissen *et al.*, 2001). Genomic selection applied in a pig breeding scheme combining information of performances from purebreds and crossbreds can significantly increase genetic gain and lower rate of inbreeding compared to the scheme using only performance from purebreds only (Dekkers, 2007). The modeling of traits expressed in purebreds and crossbreds is similar to the GxE modeling of a trait expressed in B and C environments. In another deterministic simulation study by van Grevenhof and van der Werf (2015), genomic selection was used to investigate genetic gain of breeding programs using purebred and crossbred records. The variables investigated were proportion of purebred *versus* crossbred animals in the reference population, r_g between purebred and crossbred traits and economic weight on performance of purebreds and crossbreds. The study showed that with r_g of 0.5 and 0.7, increasing proportion of crossbreds from 0-100% in the reference population increased genetic gain of the breeding program, but with an r_g of 0.9, inclusion of crossbred animals in the reference population reduced genetic gain. However, to our knowledge, no studies have explored genomic selection breeding program for sib testing of broiler chicken.

Proportion of birds phenotyped in B and C and the level of GxE interactions are important factors to consider in designing broiler chicken breeding schemes (Bijma and Arendonk, 1998; Jiang and Groen, 1999; Mulder and Bijma, 2005). Birds in C provide information on animal performance in C, but given a limited number of birds hatched in a selection round, a high proportion of birds in C would reduce selection intensity among selection candidates remaining in B. Therefore, the key to improve genetic gain of the breeding schemes is to find the best compromise between selection intensity among selection candidates and phenotypes for the target environment. GxE also has direct effects on the optimum design of the breeding schemes. The genetic correlation between B and C traits represents the magnitude of GxE interaction, but different heritability of the traits in B and C can be also an indication of GxE.

The objective of this stochastic simulation study was to compare genomic selection broiler chicken breeding schemes when all selection candidates are kept in B environment and a proportion of birds hatched are phenotyped in C environment. Three factors were investigated: 1) proportion of birds in B *versus* C, 2) different genetic correlations between traits measured in B and C (GxE) and 3) different heritability of the trait assessed in C. In addition, sensitivity simulations were carried out to investigate the effects of genotyping strategy and breeding population structure.

3.2. Methods

Breeding schemes

The breeding schemes were simulated in 3 stages comprising 1) generating a historical base population; 2) emulating previous breeding programs that used pedigree BLUP for selection; and 3) applying genomic selection in the breeding schemes with birds in B and C environments for phenotype testing. The historical population was simulated using QMSim (Sargolzaei and Schenkel, 2009) and breeding schemes for the second and third stages were simulated using the stochastic simulation program ADAM (Pedersen *et al.*, 2009).

All the breeding scenarios simulated were derived from a common base population that was created through 2 steps as described in Alemu *et al.* (2018). The first step was to create a historical population. The simulated genome consisted of 26 chromosomes which had chromosome length ranging from 5 to 195 cM with a total length of 916 cM, closely emulating the major chromosomes in chicken (Alemu *et al.*, 2018). In the first historical generation, the number of alleles for all markers and QTL was 2 with equal frequency of 0.5. The population was simulated for 950 historical generations in order to establish mutation-drift equilibrium (Meuwissen *et al.*, 2001). Over the 950 generations, the population was gradually expanded in size from 1100 to 2400 animals with equal number of individuals from both sexes. A recurrent mutation rate of 2×10^{-5} was simulated for both markers and QTL. In descendants, markers and QTL were inherited from their parents following standard principles of Mendelian inheritance allowing for recombination (Mendel, 1866). Recombination per 100 cM was sampled from a Poisson distribution with scale parameter λ =1. The position of recombinations along a chromosome were drawn from a uniform distribution. From the historical population, a base population was created, in which each individual had 40k neutral marker loci and 2k QTLs. The marker and QTL were randomly drawn from segregating loci with a minor allele frequency of at least 0.05; and they were randomly distributed along the genome. From the common base population, birds were randomly chosen to be parents of the time steps 1-5. A time step is a selection round, in which offspring is born and tested for phenotypes, and selection is carried out. In time steps 6-8, the selected birds from time steps 1-3 were mature enough to be parents, but the remaining number of parents was from the base population to make up the parental group of 16 roosters and 160 hens in a time step. From time step 9 and onwards, the parents were no longer from the base population, but they were selected birds from previous time steps. Different genetic parameters were used to simulate traits for birds at the base population.

The simulated breeding schemes had overlapping generations, and in each generation, selection was carried out at several time steps. A generation was equivalent to 6.5 time steps. At each time step, a parental group of 16 roosters and 160 hens were randomly mated to produce 1280 offspring birds with 8

offspring per hen. Parents produced offspring for several consecutive time steps. Sex ratio among the offspring was 1:1 between males and females. The breeding schemes were simulated for 40 time steps.

In the first 20 time steps, all 1280 birds hatched in each time step were phenotyped in the B environment only, and all the birds were selection candidates. Selection during this stage was based on pedigree-based BLUP EBVs estimated from records in B only. This stage was to mimic the situation of breeding programs in which broiler chicken have been selected for a certain period using records in B. All simulated breeding schemes had the same designs in the first 20 time steps.

From time step 21- 40, the 1280 birds hatched in each time step were all genotyped and allocated to either B or C environment for phenotype testing. The birds, therefore, had performance records in either B or C, and the birds in C were siblings of birds in B. The number of birds in B or C depended on the scenario of the breeding schemes. After genotyping and assessment of phenotype, single step GBLUP (ssGBLUP) models (Aguilar *et al.*, 2011; Christensen and Lund, 2010) were used in each time step to estimate GEBVs of all birds. Instead of GBLUP, ssGBLUP was used to utilize all pedigree and phenotype information of birds from the previous time steps 1-20 when genomic information was not available. Based on GEBV rankings, breeding parents were always selected from birds in B. Due to bio-security restriction, the C birds can't be used as candidates for selection.

Factors investigated

The factors investigated in this study were the genetic correlation (r_g) between trait records obtained in B and C, heritability of C trait and proportions of birds in B *versus* C (Table 3.1). The genetic factors of r_g and heritability were used for trait simulation of birds at the base population. Birds hatched at each time step 21-40 were either in B or transferred to C at different proportions for phenotype assessment. On average, selection intensity for breeding schemes with 0, 15, 30 and 45% birds transferred to C was 2.82, 2.77, 2.70 and 2.62, respectively, for males, and selection intensity was 1.97, 1.90, 1.81 and 1.69, respectively, for females.

Table 3.1: Variation of factors in the breeding programs simulated

Investigated factors	**Levels**
Proportion of birds transferred to C (%)	0; 15; 30; 45
Heritability of B trait (trait expressed in B environment)	0.28
Heritability of C trait (trait expressed in C environment)	0.15; 0.25; 0.35
Genetic correlation (r_g) between traits measured in B and C	0.5; 0.7; 0.9

In a broiler-breeding program, the overall breeding goal includes a number of traits with different economic weights. However, in this study, to simplify, only a growth performance-like trait was

considered, as it is the primary trait in the breeding goal for all broiler-breeding companies (Hiemstra and ten Napel, 2013). The simulated trait expressed in B and C environments was similar to growth performance in the two environments, and thus its genetic parameters were simulated based on parameters for growth performance in B and C from studies on broiler chicken (Kapell *et al.*, 2012; Kause *et al.*, 2012; Momen *et al.*, 2017; N'Dri *et al.*, 2007). The trait expressed in B environment was defined as B trait, and the trait expressed in C environment was defined as C trait.

Combination of the three factors: proportions of birds in C (4 levels), heritability of C trait (3 levels) and r_g between B and C traits (3 levels) yielded a total of 36 different scenarios to be simulated. Of the three investigated factors, r_g and heritability of C trait are properties of a population, which only to a limited extent can be manipulated by breeders. On the other hand, proportion of birds in C can be altered, resulting in different breeding schemes. Schemes without birds phenotyped in C included 9 scenarios, and the remaining 27 scenarios had a proportion of the birds transferred to the C environment.

Trait simulation

The traits expressed in B or C had equal means of 0 and equal genetic variances of 1 in the base population. Theoretically, GxE due to differences between B and C can result in different ranking of breeding values of birds in B and C, different heritability in the two environments, and different genetic variance in the two environments (Lynch and Walsh, 1998). In this simulation, the two first effects were accounted because GxE was modelled through r_g and heritability of C trait. However, different genetic variation was not modelled, but was assumed to be identical in B and C. This is because heterogeneity of genetic variance does not change rankings between selection candidates when the candidates are located in a single environment, their sibs are in another environment, and performances in the two environments are treated as two correlated traits (Sae-Lim *et al.*, 2013). Non-additive genetic effects were not included in simulation of the studied traits.

The phenotype of the trait expressed in B or C for the i^{th} bird in the base population, y_i, was calculated as $y_i = g_i + e_i$, where g_i is the true breeding value (TBV) of the i^{th} bird in the base population for a phenotypic record in B or C, and e_i is the environmental term for a phenotypic record in B or C. Each animal had TBVs for both B and C traits. Calculation of the TBVs were from 2000 QTLs. The effects of QTL $\begin{bmatrix}\alpha_B \\ \alpha_C\end{bmatrix}$ were scaled to achieve an initial genetic covariance matrix of $\begin{bmatrix}1 & r_g \\ r_g & 1\end{bmatrix}$ in the base population. All additive genetic variance and covariance were explained by the additive QTL variance and covariance. During simulation of breeding scenarios, effects of each QTL was kept constant, but allele frequency of each QTL might change due to selection and drift.

The environmental term for both B trait and C trait was drawn from a random normal distribution *N[0,(1-h^2)/h^2]*, where h^2 is heritability of B and C traits, respectively. Environmental variance was kept constant through the simulations, regardless of the changes in additive genetic variance. Environmental covariance between B and C traits was 0 as each bird could have phenotypic records measured in only one environment, either B or C.

Selection criteria

It was assumed that the breeding goal had an economic value of 1 for performance of birds in C and an economic value of 0 for performance of birds in B. The breeding goal was applied for selection from time step 21-40 and for assessing genetic merit of all breeding schemes. However, during the period of time steps 1-20, to emulate previous breeding program, the selection index had economic value of 1 for performance of birds in B and 0 for performance of birds in C.

Selection was based on bivariate model:

$$\begin{bmatrix}\mathbf{y_B}\\ \mathbf{y_C}\end{bmatrix} = \begin{bmatrix}\mathbf{X_B} & 0\\ 0 & \mathbf{X_C}\end{bmatrix}\begin{bmatrix}\mathbf{b_B}\\ \mathbf{b_C}\end{bmatrix} + \begin{bmatrix}\mathbf{Z_B} & 0\\ 0 & \mathbf{Z_C}\end{bmatrix}\begin{bmatrix}\mathbf{g_B}\\ \mathbf{g_C}\end{bmatrix} + \begin{bmatrix}\mathbf{e_B}\\ \mathbf{e_C}\end{bmatrix} \quad (3.1)$$

where $\mathbf{y_B}$ and $\mathbf{y_C}$ are the vectors of phenotypic records of birds in B and C; $\mathbf{b_B}$ and $\mathbf{b_C}$ are vectors of fixed effects of time step for the records in B and C; $\mathbf{g_B}$ and $\mathbf{g_C}$ are vectors of breeding values of the B and C traits; $\mathbf{X_B}$ and $\mathbf{Z_B}$ and $\mathbf{X_C}$ and $\mathbf{Z_C}$ are incidence matrices associating the fixed effects and the breeding values to the phenotypic records in B and C; $\mathbf{e_B}$ and $\mathbf{e_C}$ are vectors of random residuals in B and in C, respectively.

The model (3.1) assumed $\begin{bmatrix}\mathbf{e_B}\\ \mathbf{e_C}\end{bmatrix} \sim MVN\left[0, \begin{pmatrix}\mathbf{I_B}\sigma^2_{eB} & 0\\ 0 & \mathbf{I_C}\sigma^2_{eC}\end{pmatrix}\right]$, where $\mathbf{I_B}$ and $\mathbf{I_C}$ are identity matrices corresponding to birds in B and C environments; σ^2_{eB} and σ^2_{eC} are environmental variances of B and C traits, respectively.

For time steps 1-20, selection was based on EBVs estimated from the bivariate model (3.1) using the pedigree-based BLUP approach (Henderson, 1975) although there were no phenotypic records for C trait. In the BLUP model, the breeding values were assumed to follow a multivariate normal distribution $MVN[0, \mathbf{A} \otimes \mathbf{V_g}]$, where **A** is the matrix of additive genetic relationships based on the pedigree; $\mathbf{V_g}$ is a genetic covariance matrix of B and C traits as a 2x2 matrix; $\otimes$ is the Kronecker product. The pedigree relationship matrix was constructed from pedigree traced back to the base population.

For time steps 21-40, selection was based on GEBVs estimated from the bivariate model (3.1) using the ssGBLUP approach (Aguilar *et al.*, 2011; Christensen and Lund, 2010) for all scenarios. In the ssGBLUP model, it was assumed $\begin{bmatrix}\mathbf{g_B}\\ \mathbf{g_C}\end{bmatrix} \sim MVN[0, \mathbf{H} \otimes \mathbf{V_g}]$, where **H** is a combined matrix of pedigree relationship matrix **A** and genomic relationship matrix with weight value ω = 0.25 (Aguilar *et al.*, 2011; Christensen and

Lund, 2010) on the pedigree relationships. The genomic relationship matrix was constructed based on marker data (VanRaden, 2008).

In each time step, EBVs or GEBVs were predicted for all individuals after all records in that time step were obtained using the model above. Computations were carried out using the DMU5 module of DMU package (Madsen and Jensen, 2013). The prediction in each time step, for example time step 40, used all information (phenotypes, genomic data and pedigree) of all individuals from time step 1 till time step 40. Therefore, although all birds were genotyped in each of the time step 21-40, ssGBLUP was performed for genetic evaluation in order to utilize the phenotypic records from time step 1-20. Selection of birds to become parents were carried out right after genetic evaluation even though birds were not yet sexually mature at the time of selection.

Sensitivity analysis

A sensitivity analysis was carried out to compare genetic gain of breeding schemes that used different genotyping strategies and number offspring per hen. In the main simulation study, we assumed that all birds were genotyped when genomic selection was introduced in the breeding program. To ensure the general validity of our results, extra simulations were carried out to investigate sensitivity when not all birds were genotyped and/or the number of offspring per hen was increased.

The breeding programs used in sensitivity analysis was similar as described in the main simulation study above, but with a few modifications (Table 3.2). For all cases investigated, in the sensitivity analysis, the level of GxE were limited to only one case where r_g was 0.7 and heritability of C trait was 0.15. In each time step, only 50% of the total number of birds hatched in each time step was genotyped. Birds were selected randomly for genotyping. In sensitivity analysis simulation 1 (SS1), only 50% of birds in B and C were genotyped, and number of offspring per hen was changed. Number of offspring was 8 or 10 birds hatched from each hen in each time step. Therefore, in total, SS1 had 8 scenarios that had 1200 or 1600 birds with 0, 15, 30 and 45% birds in C. In sensitivity analysis simulation 2 (SS2), different genotyping strategies were designed for the breeding schemes with 15 and 30% birds in C. The strategies included different proportions of genotyping allocated to the birds in B *versus* C. Number of offspring per hen was 8 in each time step.

Table 3.2: Sensitivity analysis 1 and 2 (SS1 and SS2) simulating breeding schemes with varied proportion of birds in C using different genotyping strategies and total number of birds hatched for phenotype testing in each time step.

Variables investigated	**SS1**								**SS2**					
Total number of birds hatched for phenotyping	1280				1600				1280					
Total number of genotyped birds	640				800				640					
Scenario	H8-P0	H8-P15	H8-P30	H8-P45	H10-P0	H10-P15	H10-P30	H10-P45	P15-GC15	P15-GC30	P30-GC15	P30-GC30	P30-GC45	P30-GC60
Number of birds moved to C (proportion of birds hatched)	0 (0%)	192 (15%)	384 (30%)	576 (45%)	0 (0%)	240 (15%)	480 (30%)	720 (45%)	192 (15%)	192 (15%)	384 (30%)	384 (30%)	384 (30%)	384 (30%)
Number of birds in C genotyped (proportion of genotyping)	0 (0%)	96 (15%)	192 (30%)	288 (45%)	0 (0%)	120 (15%)	240 (30%)	360 (45%)	96 (15%)	192 (30%)	96 (15%)	192 (30%)	288 (45%)	384 (60%)

Simulation outputs

For each scenario, 50 replicates were simulated. For each replicate, genetic merit (G_t) at time step *t* was the average of TBVs of the C trait of all birds hatched in time step *t*. The difference between genetic merits at time step 31 (G_{31}) and 40 (G_{40}) was used to compute rate of genetic gain per time step (ΔG):

$$\Delta G = (G_{40} - G_{31})/(40 - 31) \quad (3.2)$$

The inbreeding coefficient of each individual was the proportion of homozygous identical-by-descent markers for the individual (Pedersen *et al.*, 2010). The average inbreeding coefficient F_t at time step *t* was the average of the inbreeding coefficients of the 1280 individual birds hatched at time step *t*. To be comparable to other studies, rate of inbreeding per generation was used for data analysis instead of the rate of inbreeding per time step. Therefore, in calculating inbreeding coefficient, time step *t* was translated to its corresponding generation. For each replicate, rate of inbreeding per generation (ΔF) (Liu *et al.*, 2016):

$$\Delta F\ (\%) = (1 - e^{\beta}) * 100 \quad (3.3)$$

where β is slope of the linear regression of $ln(1-F_t)$ on generation corresponding to time step 31-40.

In addition, accuracy of ssGBLUP prediction was computed as the correlation between predicted breeding values and TBVs of the C trait from all B birds hatched at time step 36 after the ssGBLUP evaluation was done for the time step. Accuracy was evaluated at time step 36 because birds selected at this time step were the last selected parents that produced offspring at time step 40.

Data analysis

The variables *ΔG* and *ΔF* for each replicate were used for comparisons among scenarios in the main simulation study while only variable of *ΔG* was used for assessing differences among scenarios in the sensitivity analysis. Descriptive statistics and standard ANOVA was used. Comparison tests for significance using Tukey's HSD (honest significant difference, P <0.05) were used. For accuracy of ssGBLUP prediction, only means were reported.

In the main study, three factors were involved in the ANOVA model including proportion of birds in C, r_g and heritability of C trait. Their main effects as well as all two and three-factor interactions were assessed. In the sensitivity simulations, each of SS1 and SS2 had 4 scenarios while SS3 had 6 scenarios. Rate of genetic gain of 8 SS1 scenarios and 4 corresponding scenarios of the main study that had the same r_g and heritability was combined for analysis of a two-way ANOVA model. One factor included in the ANOVA was number of birds genotyped and number of offspring per hen (3 levels). The other factor was proportion of birds in C (4 levels). For SS3, one-way ANOVA model was applied comparing 6 scenarios.

3.3. Results

Rate of genetic gain

Three-factor interaction between proportion of birds in C, r_g and heritability of C trait on *ΔG* was not significant (P=0.099). Significant interactions on *ΔG* were found between proportion of birds in C and r_g (P <0.001) and between the proportion and heritability (P <0.001). The interaction between r_g and heritability did not have a significant effect on *ΔG* (P =0.562). Figure 3.1 shows genetic gain of breeding schemes at different r_g between B and C traits and different heritability of trait recorded in C.

The two-way interaction between proportion of birds in C and r_g on ΔG was significant. With r_g of 0.5 and 0.7, *ΔG* of the breeding schemes without birds in C was significantly lower than schemes with birds in C. On average, the schemes without birds in C had *ΔG* of 0.116 per time step with r_g of 0.5 and 0.164 with r_g of 0.7 while the schemes with birds in C had comparable *ΔG* of 0.199 and 0.200 with r_g of 0.5 and 0.7, respectively. With r_g of 0.5 and 0.7, among the schemes with birds in C, *ΔG* of the schemes with 30% and 45% birds in C were significantly higher than *ΔG* of the scheme with 15% birds in C (P<0.05). With r_g of 0.9, *ΔG* of the schemes with 0, 15 and 30% birds in C were not significantly different between each other (P>0.05), but they were significantly higher than *ΔG* of the schemes with 45% birds in C (P<0.05).

The change of *ΔG* differed with increasing r_g when the proportion of birds in C varied. Increasing r_g significantly enlarged *ΔG* of the schemes without birds in C. The increase in *ΔG* was also observed in the scheme with 15% birds in C when r_g increased. However, *ΔG* of the scheme with 30% birds in C did not

show significant difference as r_g increased. Meanwhile, there was a decreasing trend of *ΔG* for the scheme with 45% birds in C when r_g increased. In other words, among the breeding schemes, *ΔG* of the schemes with 30% birds in C were similar at different r_g.

The interaction between proportion of birds in C and heritability was significant for *ΔG*. With heritability of 0.15 and 0.25, changing the proportion of birds in C led to significant differences in *ΔG* between schemes with birds in C. With a heritability of 0.35, the differences in *ΔG* due to proportion of birds in C were not significant between the schemes with birds in C. More importantly, as heritability increased, an increase in *ΔG* was observed in the schemes with birds in C. However, *ΔG* in the scheme without birds in C, as expected, was not affected by heritability. On average, *ΔG* of scenarios with heritability of 0.15, 0.25 and 0.35 were 0.161, 0.163 and 0.161, respectively, for the schemes without birds in C, *ΔG* were 0.191, 0,200 and 0.210, respectively, for the schemes with birds in C.

In addition, two-factor interaction between effects of proportion of birds in C and heritability on *ΔG* suggested that the schemes with birds in C had significantly higher *ΔG* than the schemes without birds in C at all levels of heritability of 0.15, 0.25 and 0.35. However, without consideration of r_g, this might give a misleading interpretation of the results. For example with r_g of 0.9, effect of heritability was not enough to cause the differences in *ΔG* among breeding schemes with 0, 15 and 30% birds in C although with heritability of 0.25 and 0.35, the schemes with 15% and 30% birds in C tended to have higher *ΔG* than the schemes without birds in C ($P>0.05$). The differences in *ΔG* among breeding schemes depended on both parameters of r_g and heritability. In addition, the p-value for the three-factor interaction was relatively low ($P=0.099$). Therefore, results of *ΔG* were presented in three-factor interaction between proportion of birds in C, r_g, and heritability of C trait (Figure 3.1).

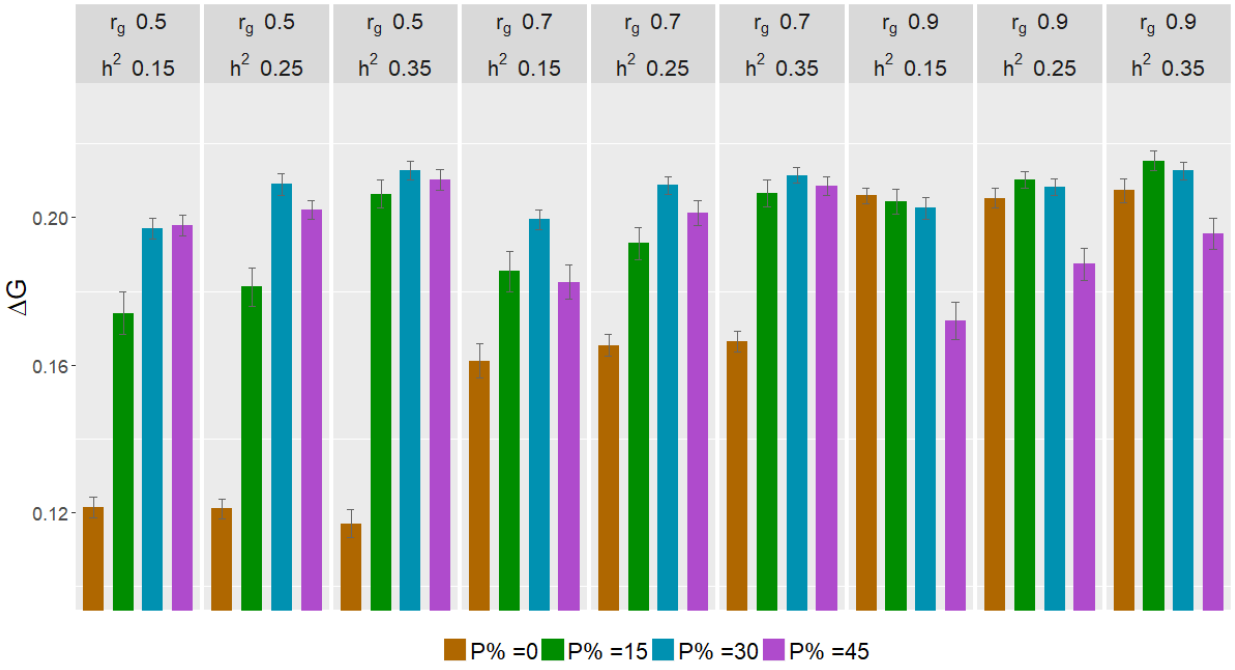


Figure 3.1: Means of rate of genetic gain per time step (*ΔG*) (±Standard errors) of scenarios with different proportion of birds in C (P%) when genetic correlation (r_g) between B and C traits and heritability (h^2) of C trait was varied.

Rate of inbreeding

The three-factor interaction on *ΔF* was not significant between proportion of birds in C, r_g and heritability of C trait (P=0.445). Significant interactions on *ΔF* were found between proportion of birds in C and r_g (P =0.005) and between r_g and heritability (P =0.043). The interaction on *ΔF* between proportion of birds in C and heritability was not significant (P =0.085).

The proportion of birds in C affected *ΔF* of the breeding scheme differently as factors of r_g changed (Figure 3.2). With increasing r_g, *ΔF* of the schemes without birds in C did not change. On the other hand, *ΔF* of the schemes with birds in C decreased with increasing r_g. On average, *ΔF* of the schemes with r_g of 0.5, 0.7 and 0.9 were 3.27, 2.99 and 2.62%, respectively. As proportion of birds in C increased, the *ΔF* reduced. With r_g of 0.5, *ΔF* of the schemes with 0 and 45% birds in C were lowest, and *ΔF* of the scheme with 15% birds in C was highest. With r_g of 0.7, *ΔF* among schemes was not significantly different. With r_g of 0.9, the schemes without birds in C had the highest *ΔF*, followed by the schemes with 15, 30 and 45% birds in C.

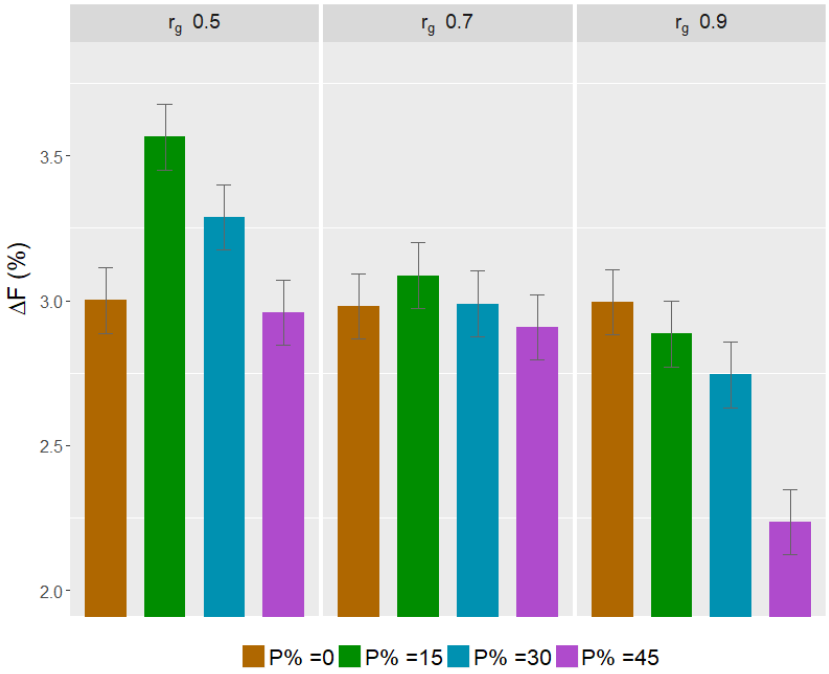


Figure 3.2: Means of rate of inbreeding per generation (*ΔF*%) (± Standard error of means) of breeding schemes with different proportion of birds in C at different genetic correlation (r_g) between B and C traits. Significant interaction on *ΔF* was found between the effects of r_g and heritability of C trait. With r_g of 0.7, *ΔF* showed no significant differences for all levels of heritability. With r_g of 0.5, *ΔF* had decreasing tendency with increasing heritability. With r_g of 0.9, the trend was opposite that *ΔF* was increasing when heritability increased.

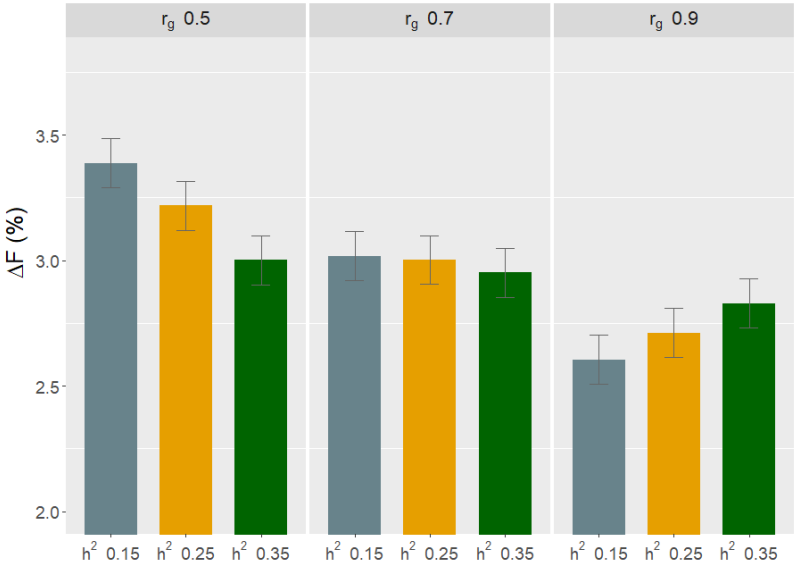


Figure 3.3: Means of rate of inbreeding per generation (*ΔF* %) (± Standard error of means) of at different genetic correlation (r_g) between B and C traits and heritability (h^2) of C trait.

Sensitivity analysis

In SS1, breeding schemes with 8 and 10 offspring per hen in each time step had 0, 15, 30 and 45% birds in C when only 50% of birds in B and C were genotyped. It was found that the schemes of SS1 had lower *ΔG* than the corresponding schemes in the main study. In SS1, the schemes with 8 offspring per hen per time step had lower *ΔG* than the schemes with 10 offspring per hen. However, similar to the main study, the schemes without birds in C had the lowest *ΔG* among breeding schemes investigated in SS1 (Figure 3.4). Also, *ΔG* of the scheme with 30% birds in C was highest followed by the schemes with 15 and 45% birds in C when number of offspring per hen was 8. Meanwhile, *ΔG* had increasing tendency as the proportion of birds in C increased from 0 to 45% when number of offspring per hen was 10. However, the rate of the increase in *ΔG* reduced with increasing proportion of C birds. The difference in *ΔG* between the schemes with 30 and 45% birds in C was relatively little.

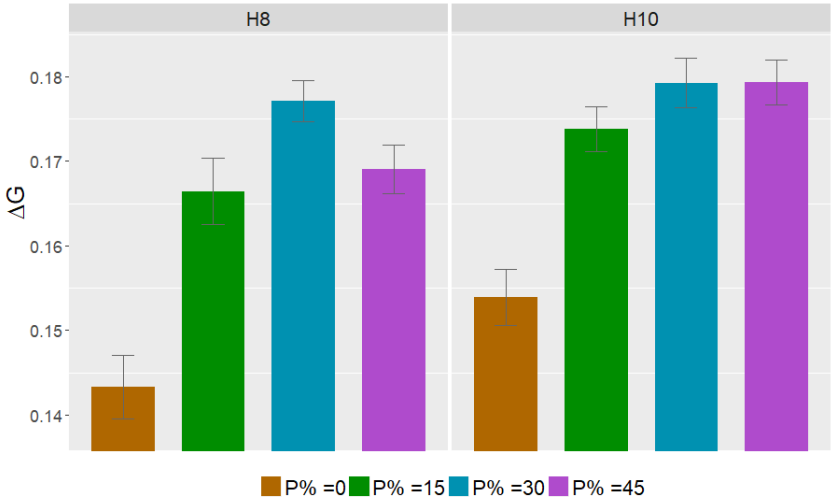


Figure 3.4: Means of rate of genetic gain per time step (*ΔG*) (±Standard errors) of sensitivity simulation 1 - SS1 for breeding schemes with 8 (H8) or 10 (H10) offspring per hen per time step and different proportion of birds in C (P%).

In SS2, genetic gain of breeding schemes with 15 and 30% birds in C that used different genotyping strategies was examined when number of genotyped birds was kept constant (Figure 3.5). For the breeding schemes with 15% birds in C, the scheme with 30% genotyping allocated to the birds in C gave higher *ΔG* than the scheme with 15% of genotyping allocated to the birds in C. For the breeding scheme with 30% birds in C, the strategy with 30% of genotyping allocated to the birds in C gave the highest *ΔG* compared to other strategies. Given a constant number of genotyped birds, there was decreasing tendency of *ΔG* as the proportion of genotyping allocated to the birds in C increased from 30% to 60%.

Among the six schemes of SS3, *ΔG* was highest in the scheme with 30% birds in C and 30% genotyping allocation to the birds in C.

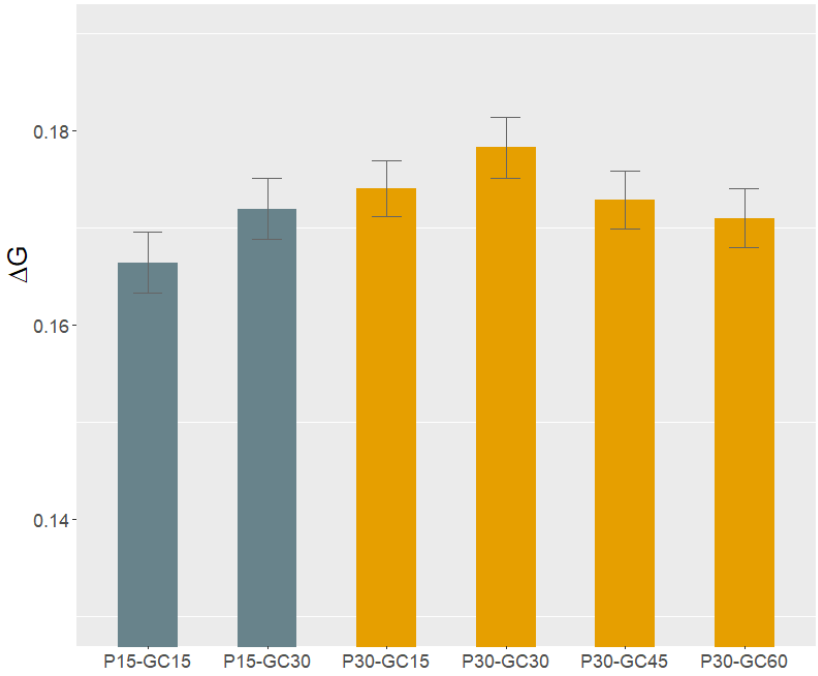


Figure 3.5: Means of rate of genetic gain per time step (*ΔG*) (±Standard errors) of sensitivity simulation 3 – SS2 for the breeding schemes with 15 and 30% (P15 and P30) birds in C using different genotyping strategies that were 15, 30, 45 and 60% of genotyping proportion allocated to the birds in C (GC15, GC30, GC45 and GC60)

3.4. Discussion

In this study, we have investigated genetic gain (*ΔG*) and rate of inbreeding (*ΔF*) in different genomic breeding schemes for broiler chicken with varying amounts of GxE between bio-secure breeding environment (B) and commercial environment (C). Schemes investigated had 0%, 15%, 30% and 45% birds in C. Effects of GxE were modelled by varying the genetic correlation (r_g) between B and C traits (0.5, 0.7 and 0.9), and heritability of C trait (0.15, 0.25 and 0.35). Sensitivity analysis was also carried out to further investigate effects of genotyping strategy and increased number of offspring per hen.

Rate of genetic gain

Genetic gain in the main study was influenced by proportion of birds in B *versus* C, r_g between B and C traits, and heritability of C trait. The proportion of birds in B *versus* C has effects on accuracy of selection and selection intensity. Since only performance in C has economic value, a higher proportion of birds in C results in higher accuracy of GEBV prediction. For example, when r_g was 0.5 and heritability was 0.25,

accuracy of GEBV prediction was 0.369, 0.718, 0.777 and 0.804 for 0, 15, 30 and 45 % birds in C, respectively. When r_g was 0.9 and heritability was 0.25, accuracy of GEBV prediction was 0.745, 0.809, 0.819 and 0.837 for 0, 15, 30 and 45 % birds in C, respectively. Even with r_g of 0.9, an increasing proportion of birds in C leads to increasing accuracy of GEBV prediction. As a consequence, genetic gain of a breeding program can be improved by introducing records in the C environment if there is significant genotype by environment interactions present in the population. This was illustrated in Bijma and Arendonk (1998) and Mulder and Bijma (2005), who found that genetic gain was improved with extra information from C when r_g was less than 1. However, given a limited hatching and reproductive capacity, increasing number of birds in C reduces selection intensity in B due to fewer selection candidates. In other words, there is a trade-off for additional accuracy of selection from records in C and a reduction in selection intensity. In our main study, 30% birds in C were the balance point between the accuracy of selection and selection intensity.

The optimum choice of a breeding scheme depends on the extent of GxE. Transferring birds to C did not improve ΔG in all situations. The level of r_g changed accuracy of selection in the scenarios investigated. For example, when proportion of C birds was 0, and heritability was 0.25, accuracy of GEBV prediction was 0.369, 0.534 and 0.745 for r_g of 0.5, 0.7 and 0.9, respectively. The contribution of records in B to accuracy of selection increased as r_g between the two environments increased. In other word, the contribution of records in C relative to the contribution of records in B to accuracy of GEBV prediction decreases as r_g increases. This explained the increase in ΔG of the schemes without C birds with increasing r_g. The difference in ΔG between the schemes with and without C birds was also smaller when r_g increased or there was less GxE. Dekkers (2007) concluded that when r_g was 0.7, genomic selection could improve genetic gain if information from records in B and C was combined. However, van Grevenhof and van der Werf (2015) implied that with r_g of 0.9, transferring animals from B to C environment did not increase the genetic gain. Therefore, when the level of GxE is low (r_g =0.9), transferring birds from B to C environment is not necessary. Nonetheless, with the low level of GxE, a possible benefit of birds tested in C is to enable selection for disease resistance traits in the C environment, especially diseases that persist in C environment, but do not exist in B environment (Morton *et al.*, 2010).

Apart from r_g, heritability of C trait had significant effect on ΔG of breeding schemes with birds in the C environment. Different levels of r_g relates to the contribution of records in B to the accuracy of GEBV prediction while different heritability of C trait relates to the contribution of records in C to the accuracy. As heritability of records in C increases, the relative information content of the records to predict additive genetic merit increases, and thus the contribution of the records to accuracy of GEBV prediction increases.

For example, when r_g was 0.7 and proportion of birds in C was 30%, accuracy of GEBV prediction was 0.756, 0.793 and 0.827 for heritability of 0.15, 0.25 and 0.35, respectively. However, in the schemes without C birds, heritability of C trait is not important. Response of these schemes depends on amount of genetic variation and r_g. In our simulation, the unit of *ΔG* is in genetic standard deviations, and genetic standard deviation is kept constant as heritability of C trait changes. Therefore, no change on *ΔG* is observed in the schemes without birds in C as heritability of C trait increases.

In designing breeding programs, effects of GxE are often modelled by including r_g while heterogeneous heritability of traits across environments is often not taken into account (Bijma and Arendonk, 1998; Dekkers, 2007; Mulder and Bijma, 2005; van Grevenhof and van der Werf, 2015). The value of r_g expresses the magnitude of GxE interaction, but different heritability of B and C traits can be also a consequence of the GxE interaction. Heritability of C trait can be lower or higher than heritability of B trait. In the study by Kapell *et al.* (2012), in three out of four studied pure broiler lines, the heritability of C trait was lower than B trait for body weight at 5 weeks of age. The higher and lower heritability of C trait than B trait was also found in the study by N'Dri *et al.* (2007). In addition, different heritability of B trait can have significant effect on designing GxE breeding program, but this is not included in our study. The contribution of records in B would increase with increasing heritability of B trait.

In addition to proportion of birds in C, r_g and heritability of C trait, the sensitivity analysis showed that amount of genotyping, number of offspring per hen and genotyping strategies can influence genetic gain of a genomic selection program for broilers. In the sensitivity analysis, only 50% of all birds hatched were genotyped, r_g was 0.7, and heritability of C trait was 0.15. Genetic gain of the corresponding scenarios in the main study was higher than in the SS1 scenarios. The higher genetic gain was found even when SS1 scenarios had 10 offspring per hen and the scenarios in the main study had 8 offspring per hen hatched in each time step. This is primarily due to the higher number of birds genotyped in the main study than in the SS1 scenarios. The main study has 1280 birds genotyped in each time step while SS1 has 640 and 800 birds genotyped for the schemes with 8 and 10 offspring per hen, respectively.

When number of offspring per hen was 8, the relative differences between the SS1 schemes with 0, 15, 30 and 45% birds in C had a similar tendency to those of the main study. The schemes with 30% birds in C had the highest *ΔG*. In this case, selection intensity was not changed, and proportions of birds in B *versus* C among the schemes were the same as the proportions of birds in B *versus* C genotyped. When the number of offspring per hen increased from 8 to 10 per time step, the schemes with 45% birds in C had the highest *ΔG*. However, the difference between the schemes with 30 and 45% was very small. This

implies that the 30% birds in C is close to the optimum proportion when number of offspring per hen is 8 and 10 per time step.

In SS2, *ΔG* of the scheme with 15% birds in C tended to increase as number of genotyped birds in C increased. This result suggests that when proportion of birds in C is lower than the optimum proportion, higher proportion of birds in C should be genotyped. Therefore, the scheme that 15% of birds were tested in C for phenotyping and 100% of these birds were genotyped was expected to have higher *ΔG* than the scheme that 30% of birds were tested in C and 50% of these birds were genotyped. However, the lower *ΔG* of the scheme with 15% birds in C than the scheme with 30% birds in C could not be offset by increasing number of genotyped birds in C and increased selection intensity. The scheme with 30% of birds in C and 30% genotyping allocation to the birds in C has the highest *ΔG* in SS2. Increasing number of genotyped birds in C increases amount of information from C environment. However, when genotyping resources are limited, the increase in number of C birds genotyped reduces number of B birds genotyped. Information "brought" from C to B environment is less efficient with fewer number of B birds genotyped. This may be an explanation for the decreasing tendency of *ΔG* as the proportion of genotyping allocated to the birds in C increased from 30% to 60%. The result of SS2 implies that to maximize genetic gain of a breeding program for commercial broiler chicken with limited genotyping effort, the optimum proportion of birds allocated to B and C environment should be the optimum proportion of genotyping allocated to B and C birds, respectively.

A genotyping strategy, which was not tested in sensitivity analysis, was selective genotyping. Boligon *et al.* (2012) found that the selective genotyping strategy improves predictive ability of breeding values, and that animals with best performance are the most informative. This is possible in broilers where important traits like body weight and feed efficiency can be obtained before sexual maturity. Alemu *et al.* (2018) also illustrated an increase in genetic gain with selective genotyping in genomic selection programs for broilers. In applying this strategy, it should be considered whether selective genotyping should be applied for birds in B, C or both of the environments. Selective genotyping especially in B can be advantageous in order to increase genetic gain for a given investment in genotyping.

Rate of inbreeding

Along with genetic gain, inbreeding of the schemes in the main study was investigated. It was shown that the proportion of birds in C, r_g and the heritability of C trait all affected *ΔF*. Among the schemes utilizing records from C, *ΔF* decreased as r_g increased or proportion of birds in C increased. Heritability of C trait had different effects on *ΔF* when r_g varied.

Transferring birds from B to C environment reduces selection intensity and increases the amount of information from C. Reducing selection intensity reduces *ΔF* because decreasing the number of selection candidates decreases probability of co-selecting birds from the same family. Increasing the amount of information from C has two opposite consequences on *ΔF*. One consequence leads to an increased probability of co-selecting close relatives due to utilizing information from C, thus increasing *ΔF*. Utilization of information from C causes inbreeding. For example, a group of close relatives receives similar information from C, and therefore probability of co-selecting the close relatives increases. In addition, increasing proportion of birds in C increases weight or the contribution of C information to prediction of GEBVs of birds in B, which increases *ΔF*. At the same time, another consequence of increased the amount of information from C is an increase in accuracy of prediction, especially when genomic information describes relationships between full-sibs better than pedigree information. As amount of information from C increases, accuracy of GEBVs of birds in B increases, and therefore probability of co-selecting close relatives due to utilizing information from C decreases.

An extra simulation was carried out to test the consequences of *ΔF* when information from C increased and selection intensity remained constant. Heritability of C trait was 0.15, and r_g was 0.5 and 0.9. The breeding scheme for this simulation was same as in main simulation, except that number of offspring per hen was varied and number of birds in B was kept constant. Number of offspring per hen was 4, 5, 6, 7 and 8 equivalent to 640, 800, 960, 1120 and 1280 birds hatched in each time step. In each time step, 640 birds were in B as selection candidates, the remaining was transferred to C. With r_g of 0.9, *ΔF* was 2.18, 2.37, 2.49, 2.52 and 2.65 for the schemes with 4, 5, 6, 7 and 8 offspring per hen, respectively. With r_g of 0.5, *ΔF* was 2.48, 3.29, 3.15, 2.99 and 2.82 for the schemes with 4, 5, 6, 7 and 8 offspring per hen, respectively. These results confirm that increasing information from C have two opposite consequences on *ΔF* as explained above.

When r_g is 0.5, the effect of the co-selection due to utilizing information from C is substantial on *ΔF* for the scheme with 15 and 30% birds in C. This leads to the higher *ΔF* of the schemes with 15 and 30% birds in C than the schemes without birds in C although the schemes without birds in C, indeed, has the highest selection intensity. When r_g increases, information from B environment has higher weight for the GEBV prediction and consequently reducing the co-selection due to utilizing C information. Therefore, the change of r_g relates to the change of the co-selection due to utilizing C information. This explains the reduction in *ΔF* of the schemes with birds in C as r_g increases. Meanwhile, *ΔF* of the schemes without birds in C is unaffected by the change of r_g.

When heritability of C trait increased, the changing pattern of *ΔF* were changed at different r_g. This is because the change of heritability of C trait has two opposite consequences on *ΔF*. One consequence relates to the BLUP attributes that increase in heritability decreases weight of information from relatives, thus reduces co-selection of relatives (Bijma and Woolliams, 2000; Verrier *et al.*, 1993). This reduces *ΔF*. At the same time, another consequence relates to the utilization of information from C. Increasing heritability increases weight of C information in the ssGBLUP prediction, and thus *ΔF* increases. With low r_g of 0.5, the increase in the weight of C information does not clearly show its effect while with high r_g of 0.9, it increases *ΔF*.

3.5. Conclusion

Genetic gain and rate of inbreeding of different genomic breeding schemes for broiler chicken were compared in situations with different degree of GxE between breeding (B) and commercial (C) environments. It is shown that the proportion of birds in B *versus* C for a breeding program depends on the genetic correlation between the trait assessed in B and in C (r_g), heritability of the trait measured in C, number of offspring per hen, amount of genotyping, and genotyping strategy. When r_g was 0.5 and 0.7, transferring birds to C environment increased genetic gain for the breeding program and 30% birds assessed in C was the optimum proportion. When proportion of birds in C was at the optimum proportion (30%) and genotyping efforts was limited, 30% of the genotyping effort allocated to C birds was also the optimum genotyping strategy. When proportion of birds in C was lower than the optimum proportion, genotyping more birds in C increased genetic gain in the breeding program. For rate of inbreeding, increasing the proportion of birds in C lowered inbreeding. Rate of inbreeding of the schemes with birds in C increased when r_g increased. The rate of inbreeding in schemes without birds in C did not change as r_g increased. In summary, if there is a strong GxE interaction (r_g of 0.5 and 0.7), a genomic selection scheme that uses a considerable proportion (30%) of birds to be transferred to C for phenotype testing has larger genetic gain than if all birds are tested in B environment only. In addition, rate of inbreeding is reduced as proportion of birds transferred to C increases from 15 to 45%.

References:

Aguilar, I, Misztal, I, Legarra, A & Tsuruta, S 2011, 'Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation', *Journal of Animal Breeding and Genetics,* vol. 128, no. 6, pp. 422-428.

Alemu, SW, Madsen, P, Wang, L, Henshall, J, Hawken, R, Sørensen, AC & Jensen, J 2018, 'Choice of Genotyping Strategy in Genomic Selection Programs for Broilers', *Manuscript submitted for publication*, no.

Bijma, P & Arendonk, JAM 1998, 'Maximizing genetic gain for the sire line of a crossbreeding scheme utilizing both purebred and crossbred information', *Animal Science,* vol. 66, no. 2, pp. 529-542.

Bijma, P & Woolliams, JA 2000, 'Prediction of Rates of Inbreeding in Populations Selected on Best Linear Unbiased Prediction of Breeding Value', *Genetics,* vol. 156, no. 1, pp. 361.

Boligon, AA, Long, N, Albuquerque, LG, Weigel, KA, Gianola, D & Rosa, GJM 2012, 'Comparison of selective genotyping strategies for prediction of breeding values in a population undergoing selection', *Journal of Animal Science,* vol. 90, no. 13, pp. 4716-4722.

Christensen, OF & Lund, MS 2010, 'Genomic prediction when some animals are not genotyped', *Genetics Selection Evolution,* vol. 42, no. 1, pp. 2.

Daetwyler, HD, Villanueva, B, Bijma, P & Woolliams, JA 2007, 'Inbreeding in genome-wide selection', *J Anim Breed Genet,* vol. 124, no. 6, pp. 369–376.

Dekkers, JCM 2007, 'Marker-assisted selection for commercial crossbred performance', *J Anim Sci,* vol. 85, no. 9, pp. 2104-2114.

Hayes, BJ, Visscher, PM & Goddard, ME 2009, 'Increased accuracy of artificial selection by using the realized relationship matrix', *Genet Res,* vol. 91, no.

Henderson, CR 1975, 'Best linear unbiased estimation and prediction under a selection model', *Biometrics,* vol. 31, no. 2, pp. 423-47.

Hiemstra, SJ & ten Napel, J 2013, 'Study of the impact of genetic selection on the welfare of chickens bred and kept for meat production', Final report of a project commissioned by the European Commission (DG SANCO 2011/12254).

Jiang, BX & Groen, A 1999, 'Combined crossbred and purebred selection for reproduction traits in a broiler dam line', *Journal of Animal Breeding and Genetics,* vol. 116, no. 2, pp. 111-125.

Kapell, DN, Hill, WG, Neeteson, AM, McAdam, J, Koerhuis, AN & Avendano, S 2012, 'Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments', *Poultry Science,* vol. 91, no. 3, pp. 565-74.

Kause, A, van Dalen, S & Bovenhuis, H 2012, 'Genetics of Ascites Resistance and Tolerance in Chicken: A Random Regression Approach', *G3: Genes|Genomes|Genetics,* vol. 2, no. 5, pp. 527-535.

Liu, H, Henryon, M & Sorensen, AC 2016, 'Mating strategies with genomic information reduce rates of inbreeding in animal breeding schemes without compromising genetic gain', *Animal*, no. 1-9.

Long, N, Gianola, D, Rosa, GJ, Weigel, KA & Avendano, S 2008, 'Marker-assisted assessment of genotype by environment interaction: a case study of single nucleotide polymorphism-mortality association in broilers in two hygiene environments', *J Anim Sci,* vol. 86, no. 12, pp. 3358-66.

Lynch, M & Walsh, B 1998, 'Genetics and analysis of quantitative traits', Sinauer Associates Inc., Sunderland, MA.

Madsen, P & Jensen, J 2013, 'DMU: a user's guide'. "*A package for analysing multivariate mixed models, Version 6, release 5.2. Home page: http://dmu.agrsci.dk/*",

Mendel, G 1866, 'Experiments in plant hybridization', *Verhandlungen des naturforschenden Vereins Brünn. Available online: www.mendelweb.org/Mendel.html (accessed on 1 January 2013)*, no.

Meuwissen, THE, Hayes, BJ & Goddard, ME 2001, 'Prediction of total genetic value using genome-wide dense marker maps', *Genetics,* vol. 157, no. 4, pp. 1819-29.

Michiels, T, Welby, S, Vanrobaeys, M, Quinet, C, Rouffaer, L, Lens, L, Martel, A & Butaye, P 2016, 'Prevalence of Mycoplasma gallisepticum and Mycoplasma synoviae in commercial poultry, racing pigeons and wild birds in Belgium', *Avian Pathol,* vol. 45, no. 2, pp. 244-52.

Momen, M, Mehrgardi, AA, Sheikhy, A, Esmailizadeh, A, Fozi, MA, Kranis, A, Valente, BD, Rosa, GJM & Gianola, D 2017, 'A predictive assessment of genetic correlations between traits in chickens using markers', *Genetics Selection Evolution,* vol. 49, no. 1, pp. 16.

Morton, D, Oltenacu, T, Arnould, C, Collins, L, Le Bihan-Duval, E, Hocking, P, Sørensen, P, Hartun, J, Keeling, L & Banos, G 2010, 'Scientific opinion on the influence of genetic parameters on the welfare and the resistance to stress of commercial broilers', Efsa Scientific Colloquium Summary Report, pp.

Mulder, H & Bijma, P 2005, 'Effects of genotype× environment interaction on genetic gain in breeding programs', *Journal of Animal Science,* vol. 83, no. 1, pp. 49-61.

N'Dri, AL, Sellier, N, Tixier-Boichard, M, Beaumont, C & Mignon-Grasteau, S 2007, 'Genotype by environment interactions in relation to growth traits in slow growing chickens', *Genetics Selection Evolution,* vol. 39, no. 5, pp. 513-528.

Pedersen, L, Sørensen, A & Berg, P 2010, 'Marker-assisted selection reduces expected inbreeding but can result in large effects of hitchhiking', *Journal of Animal Breeding and Genetics,* vol. 127, no. 3, pp. 189-198.

Pedersen, L, Sørensen, A, Henryon, M, Ansari-Mahyari, S & Berg, P 2009, 'ADAM: A computer program to simulate selective breeding schemes for animals', *Livestock Science,* vol. 121, no. 2, pp. 343-344.

Sae-Lim, P, Kause, A, Mulder, HA, Martin, KE, Barfoot, AJ, Parsons, JE, Davidson, J, Rexroad, CEr, van Arendonk, JA & Komen, H 2013, 'Genotype-by-environment interaction of growth traits in rainbow trout (Oncorhynchus mykiss): a continental scale study', *J Anim Sci,* vol. 91, no. 12, pp. 5572-81.

Sargolzaei, M & Schenkel, FS 2009, 'QMSim: a large-scale genome simulator for livestock', *Bioinformatics,* vol. 25, no. 5, pp. 680-1.

van Grevenhof, IE & van der Werf, JH 2015, 'Design of reference populations for genomic selection in crossbreeding programs', *Genetics Selection Evolution,* vol. 47, no. 1, pp. 1-9.

VanRaden, PM 2008, 'Efficient methods to compute genomic predictions', *J Dairy Sci,* vol. 91, no.

Verrier, E, Colleau, JJ & Foulley, JL 1993, 'Long-term effects of selection based on the animal model BLUP in a finite population', *Theor Appl Genet,* vol. 87, no. 4, pp. 446-54.

Ye, X, Avendano, S, Dekkers, JC & Lamont, SJ 2006, 'Association of twelve immune-related genes with performance of three broiler lines in two different hygiene environments', *Poult Sci,* vol. 85, no. 9, pp. 1555-69.

4

Genetic analysis on body weight at different ages in broiler chicken raised in commercial environment

Thinh T. Chu[1,2*], Per Madsen[1], Elise Norberg[1,3], Lei Wang[1], Danye Marois[4], John Henshall[4], Just Jensen[1]

[1] *Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark*

[2] *Wageningen University & Research Animal Breeding and Genomics, 6709 PG Wageningen, The Netherlands*

[3] *Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway*

[4] *Cobb-Vantress Inc., Siloam Springs, Arkansas 72761-1030, United States of America*

Corresponding author: Thinh Tuan Chu. Email: chu.thinh@mbg.au.dk

Journal of Animal Breeding and Genetics: Accepted

DOI: 10.1111/jbg.12448

Abstract

A multivariate model was developed and used to estimate genetic parameters of body weight (BW) at 1-6 week of age of broilers raised in a commercial environment. The development of model was based on the predictive ability of breeding values evaluated from a cross-validation procedure that relied on half-sib correlation. The multivariate model accounted for heterogeneous variances between sexes through standardization applied to male and female BW differently. It was found that the direct additive genetic, permanent environmental maternal and residual variances for BW increased drastically as broilers aged. The drastic increase in variances over weeks of age were mainly due to scaling effects. Ratio of the permanent environmental maternal variance to phenotypic variance decreased gradually with increasing age. Heritability of BW traits ranged from 0.28 to 0.33 at different weeks of age. The direct genetic effects on consecutive weekly BWs had high genetic correlations (0.85-0.99), but the genetic correlations between early and late BWs were low (0.32-0.57). The difference in variance components between sexes increased with increasing age. In conclusion, the permanent environmental maternal effect on broiler chicken BW decreased with increasing age from week 1-6. Potential bias of the model that considered identical variances for sexes could be reduced when heterogeneous variances between sexes are accounted for in the model.

4.1. Introduction

Optimal modelling plays a key role in improving accuracy of predicted breeding values for traits of interest, and thereby increasing genetic gain in breeding programs. Apart from direct additive genetic effects, other factors including maternal effects, sex by genotype interaction, and heterogeneous residual variances have been accounted for when modelling body weight (BW) in poultry, (Barbieri *et al.*, 2015; Begli *et al.*, 2016; Grosso *et al.*, 2010; Maniatis *et al.*, 2013; Mebratie *et al.*, 2017). However, the method used for including those factors in BW modelling varies across studies. In addition, when analyzing longitudinal data for BW in chicken, it was found that the effects of maternal factors and sex by genotype interaction change with age of the birds (Begli *et al.*, 2016; Mebratie *et al.*, 2017; Norris and Ngambi, 2006; Rovadoscki *et al.*, 2016).

Maternal effects are often ignored for egg production traits that are expressed relatively late in life, but the effects are considered to be important for growth traits in broiler chicken, especially at juvenile stage that broilers are usually selected in breeding programs (Besbes and Ducrocq, 2003). Two commonly modelled maternal effects for BW traits are additive genetic and permanent environmental effects of the

dam (Barbieri *et al.*, 2015; Maniatis *et al.*, 2013; Norris and Ngambi, 2006). The maternal effects reduce as chicken are older (Barbieri *et al.*, 2015; Maniatis *et al.*, 2013; Norris and Ngambi, 2006). Another maternal effect, age of hen, may also affect BW because strong correlations between hen age and egg weight and strong correlations between egg weight and broiler BW especially at an early age were found (Di Masso *et al.*, 1998; Tahir *et al.*, 2011; Whiting and Pesti, 1984). Failing to account for maternal effects results in reduced accuracy of selection, overestimated direct additive genetic effect and thus loss of selection response (Roehe and Kennedy, 1993).

Sex by genotype interaction for BW in commercial broiler chicken has been reported (Mebratie *et al.*, 2017). The authors suggested that modelling BW of males and females as two separate, but correlated traits could improve accuracy of selection compared to the model that assumed an average heritability and a genetic correlation of unity between BW expressed in different sexes. Moreover, the residual variance in males was larger than the variance among females for BW. The authors also found an increasing magnitude of sex by genotype interaction with increasing age. In addition, a drastic increase in residual variances was found for BW traits as chicken aged. The increasing residual variances over different ages of birds were also found in other studies (Begli *et al.*, 2016; Dana *et al.*, 2011; Mebratie *et al.*, 2017).

Different models including univariate, multivariate, and univariate random regression models have been used to analyze longitudinal data to model the development of BW in chicken over ages (Begli *et al.*, 2016; Dana *et al.*, 2011; Niknafs *et al.*, 2012; Rovadoscki *et al.*, 2016). However, in the longitudinal data of these studies, there were relatively low number of birds or few age points recorded for BW. For example, Begli *et al.* (2016) had weekly BW recorded at week 2-10, but there were only 450 birds in the experiment. With such low number of birds, some effects may not be detected by the model. Meanwhile, Mebratie *et al.* (2017) had about 646,000 birds, but BW records were at only 3 different ages of *t-7*, *t-4* and *t* days for three different groups of birds. With a low number of age points, the development of BW cannot be accurately modelled.

In addition, the production environment where birds are tested may influence BW and thus how it should be modelled. Few studies (Kapell *et al.*, 2012) present genetic parameters for broiler BW in a commercial environment. Most of the studies report the parameter for poultry BW in the breeding environment (Grosso *et al.*, 2010; Koerhuis and Thompson, 1997; Maniatis *et al.*, 2013; Niknafs *et al.*, 2012) or in controlled experimental environment (Begli *et al.*, 2016; Norris and Ngambi, 2006; Rovadoscki *et al.*, 2016). Sib-testing in both the breeding and the commercial environment is a common approach in broiler breeding programs when genotype-by-environment interactions are important. Growth performances of

birds in the commercial environment are often found to be lower than the performances of birds in the breeding environment, whereas standard deviation of BW was higher in the commercial environment (Kapell *et al.*, 2012; N'Dri *et al.*, 2007). In addition, both the absolute and the relative differences in BW between sexes are smaller in the commercial environment compared to the breeding environment (Kapell *et al.*, 2012; N'Dri *et al.*, 2007). Heritability estimates of BW in the commercial environment can be both higher or lower than the heritability of BW in the breeding environment (Kapell *et al.*, 2012; N'Dri *et al.*, 2007). Furthermore, an important difference between modelling BW traits for the breeding and commercial environments is that birds in the commercial environment provide information for selection, but they are not selection candidates because of bio-security restrictions, and thereby dams and sires of those birds do not have BW performances in this commercial environment. Proper account of selection of parents may necessitate the use of multivariate joint modeling of both commercial and breeding environments. Development of a multivariate model for BW of broiler in the commercial environment is one of the steps in building the joint model.

The current study investigated genetic parameters of BW in broilers at different ages raised in a commercial environment. The specific objectives were to (i) develop statistical models to improve accuracy of predicting breeding values, and (ii) estimate parameters for male and female BW at different ages when broilers were reared under commercial conditions.

4.2. Materials and methods

Data

Longitudinal data on BW were obtained from Cobb-Vantress who reared broiler chicken in a standard commercial production environment and recorded BW weekly. The broilers were from a purebred line primarily selected for BW in the breeding environment with very stringent regulations for bio-security. The data included 12 broiler flocks sourced from around two generations of breeder flocks. Birds in broiler flocks 1-6 had weekly records of BW from 1-6 weeks of age. Birds in broiler flocks 7-12 has weekly records of BW from 1-5 weeks of age only. Birds with unidentified sex and missing information were removed from the data. Duplicated records on the same week and BW records of 0 were also removed. Only records of BW at each week that were within four standard deviations of the mean were kept. As a result, 2.58% of birds and 0.71% of BW records were removed from the original data.

After data trimming, the number of birds was 17,967 with a total of 91,846 BW records. Birds were offspring from 253 sires and 1,187 dams. The sires and dams did not have phenotypic records in this commercial environment. In each broiler flock, all birds were hatched at the same time, sourced from

multiple breeder flocks in which dams had different ages. The mating system was hierarchical with each sire mated with multiple dams, but each dam mated with a single sire only. The pedigree covered around three generations back from the youngest birds and comprised 20,509 birds.

Development of statistical models for body weights in broilers

To develop a statistical model for BW at weeks 1-6, initially univariate models were developed to identify fixed and random factors affecting BW at different weeks of age and different sexes. The longitudinal data were divided by weeks of age and sexes into 12 subsets. For each of the twelve datasets, a univariate model was developed through removal of factors in a hierarchical fashion. A factor or interaction was removed by comparing the model with and without the effect based on criteria of model convergence, log-likelihood ratio test and predictive ability of breeding values from cross-validation as described below. The starting model included fixed effects (flock of birds, source of flocks, hatch of dam within source of flock, and age of dam when offspring were hatched) and all their possible interactions including a fourth order interaction. The random factors of the starting model were the direct additive genetic effect, maternal additive genetic effect and permanent environmental maternal effect. The final model was selected when the removal of a factor from the model significantly reduced the fit of the model or predictive ability of the model in the cross-validation. For the 12 datasets, 7 different models were selected as results of model development process based on sets of model selection criteria:

Convergence of model: The REML module (DMUAI) from the DMU software package (Madsen and Jensen, 2013) was used to estimate variance components in all models. Strict criteria for the convergence of a model were set, in which the Frobenius norms of both the update vector and the gradient vector must be lower than 10^{-5} (Madsen and Jensen, 2013).

Log-likelihood ratio test: Log-likelihood ratio tests were carried out to identify the significance of a random effect in a model by comparing models with or without the effect (significant difference, $P<0.05$).

Cross validation: Predictive ability of the univariate models was compared using cross validation. The full dataset of BW records at each week for each sex was divided into training and validation datasets based on flocks and full-sib relationships. The training dataset included all bird records from the first half of all flocks and about half of the records from the latter half of all flocks. The validation dataset included records of the other half from the latter half of all flocks. The approximately equal division of records into the training and validation datasets in the latter half of all flocks was carried out within full-sib groups.

The full dataset was used to estimate variance components of the model and to compute phenotypes corrected for fixed effects in the validation datasets using the DMUAI procedure of the DMU package. In other words, the corrected phenotype (y_c) was equal to the sum of breeding values (EBV_f), random

maternal effects and residuals estimated from the full dataset. The training dataset was used to predict breeding values of birds (EBV_v) in the validation datasets using the variance components estimated from the full dataset.

In the conventional method as used by Christensen *et al.* (2012), validation was based on correlation between y_c and EBV_v of the same individuals in the validation dataset. However, in the presence of maternal effects, this correlation might be overestimated because information from full-sibs influences EBV_v, and thus maternal effects are confounded into EBV_v. Our validation used the correlation between y_c and EBV_v of two different birds that were half-sibs because half-sibs did not share the maternal effects. Random pairs of two half-sibs were sampled for each sire, and all the pairs were used to calculate the correlation between y_c and EBV_v of two half-sibs. From the validation dataset, 6000 replicates of the pairing sampling were carried out. The number of replicates of the pairing sampling was to ensure a high probability of all birds in the validation dataset contributed to the calculation of the half-sib correlation. Predictive ability of breeding values for a model was based on the average of correlations between y_c and EBV_v ($Cor(EBV_{vi}, y_{cj})$) of two half-sib birds *i* and *j* in the validation dataset:

$$Predictive\ ability\ = \frac{Cor(EBV_{vi}, y_{cj})}{\sqrt{h^2 * r_{i,j}}}$$

Where h^2 is heritability calculated from variance components estimated from the full dataset, $r_{i,j}$ is the is the additive genetic relationship between half-sibs that is equal to 0.25.

There were seven different univariate models selected from the model development process, and the maternal additive genetic effect was included in the model only for male BW1. However, for simplification and convergence ease, consistent fixed effects across BW at different weeks of age were chosen, namely flock of birds, source of flocks and age of dam when offspring were hatched. The random effects were direct additive genetic effect, permanent environmental maternal effect and a residual.

Statistical models

For BW at each week of age, sex by genotype interaction was tested using bivariate models that treated male and female BWs as two different traits. However, it was found that the bivariate models lead to convergence failure because genetic correlations converges towards the edges of the parameter space of 1.0. Male and female BW tended to might have different variances, particularly at the later ages that selection occurs, but their correlations were not significantly different from unity. This refers to scaling effects between male and female BW. If the difference in variances between sexes was not accounted in the model, bias of prediction would increase (Thompson, 2008) and re-ranking of EBVs could occur (van der Heide et al., 2016). It was found that correlations between BW at two consecutive weeks were high,

and multivariate models that used BW1-6 as phenotypic records failed to converge. In preliminary analysis, when Legendre polynomial functions with order of two or more were fitted, the models failed to converge. With the results from univariate models for BW traits, the linear function may not describe well the change of the direct additive genetic effect over weeks of age. In addition, between two model approaches, we preferred the multivariate model to the random regression model because with more parameters, a multivariate model describes better the covariance matrix of the direct additive genetic effect.

To model male and female BW1-6, we used a multivariate model that used BW1-2 and weekly weight gains as phenotypic records, and standardized phenotypic records differently within sex and week. Weekly weight gains were calculated as the difference of BW between week 2 and 3 (WG3), week 3 and 4 (WG4), week 4 and 5 (WG5) and week 5 and 6 (WG6). The use of BW1-2 and WG3-6 in replacement for BW1-6 as phenotypic records aimed to improve the convergence of the multivariate model. The choice of these BW and weight gains was to minimize missing records due to the use of weight gains. Male and female phenotypic records were standardized separately using corresponding phenotypic standard deviations that were estimated from univariate model (4.1) for BW1-2 and WG3-5, and univariate model (4.2) for WG6:

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Za}+\mathbf{Wc}+\mathbf{e} \quad (4.1)$$

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Za}+\mathbf{e} \quad (4.2)$$

where **y** is vector of male or female phenotypic records of BW1-2 and WG3-6 at normal scale; **b** is vectors of fixed effects of flock of birds, source of flocks and age of dam. Matrices of **X**, **Z**, and **W** are incidence matrices. Vectors **a**, **c** and **e** are the direct additive genetic effect, permanent environmental maternal effect and residual, respectively. These random effects were assumed to be normally distributed: $\mathbf{a}\sim \mathbf{N}[\mathbf{0}, \mathbf{A}\sigma_a^2]$, $\mathbf{c}\sim \mathbf{N}[\mathbf{0}, \mathbf{I_d}\sigma_c^2]$ and $\mathbf{e}\sim \mathbf{N}[\mathbf{0}, \mathbf{I}\sigma_e^2]$, where **A** is the pedigree relationship matrix; $\mathbf{I_d}$ is the identity matrix for dams; **I** is the identity matrix for individual birds; σ_a^2, σ_c^2 and σ_e^2 are variances at normal scale.

Phenotypic records of males and females at each week were standardized by dividing the phenotypic records to the corresponding phenotypic standard deviations that were estimated from models (4.1) and (4.2). In the multivariate model, reduced ranks were applied to male and female traits for additive genetic effects and permanent environmental maternal effects. The standardization of phenotypic records and reduced ranks between sexes were to model the differences in variances between male and female traits and their unity correlation. The permanent environmental maternal effect of WG6 was removed due to

its insignificance. In matrix notation, the multivariate model (4.3) using standardized phenotypic records of BW1-2 and WG3-6 was:

$$\begin{bmatrix}\mathbf{y}_{1-5}^{m0}\\ \mathbf{y}_{1-5}^{f0}\end{bmatrix}=\begin{bmatrix}\mathbf{X}_{1-5}^{m} & 0\\ 0 & \mathbf{X}_{1-5}^{f}\end{bmatrix}\begin{bmatrix}\mathbf{b}_{1-5}^{m}\\ \mathbf{b}_{1-5}^{f}\end{bmatrix}+\begin{bmatrix}\mathbf{Z}_{1-5}^{m} & 0\\ 0 & \mathbf{Z}_{1-5}^{f}\end{bmatrix}\boldsymbol{a}_{1-5}+\begin{bmatrix}\mathbf{W}_{1-5}^{m} & 0\\ 0 & \mathbf{W}_{1-5}^{f}\end{bmatrix}\boldsymbol{c}_{1-5}+\begin{bmatrix}\mathbf{e}_{1-5}^{m}\\ \mathbf{e}_{1-5}^{f}\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{y}_{6}^{m0}\\ \mathbf{y}_{6}^{f0}\end{bmatrix}=\begin{bmatrix}\mathbf{X}_{6}^{m} & 0\\ 0 & \mathbf{X}_{6}^{f}\end{bmatrix}\begin{bmatrix}\mathbf{b}_{6}^{m}\\ \mathbf{b}_{6}^{f}\end{bmatrix}+\begin{bmatrix}\mathbf{Z}_{6}^{m} & 0\\ 0 & \mathbf{Z}_{6}^{f}\end{bmatrix}\boldsymbol{a}_{6}+\begin{bmatrix}\mathbf{e}_{6}^{m}\\ \mathbf{e}_{6}^{f}\end{bmatrix} \quad (4.3)$$

where $\mathbf{y}_{1\text{-}5}^{m0}$ and $\mathbf{y}_{1\text{-}5}^{f0}$ are the vectors of male and female standardized phenotypic records, respectively, for BW1-2 and WG3-5; $\mathbf{b}_{1\text{-}5}^{m}$ and $\mathbf{b}_{1\text{-}5}^{f}$ are vectors of fixed effects of bird flock, source of flocks and age of dam for males and females, respectively, of BW1-2 and WG3-5. Similarly, $\mathbf{y}_{6}^{m0}$ and $\mathbf{y}_{6}^{f0}$ are the vectors of male and female standardized phenotypic records, respectively, for WG6; $\mathbf{b}_{6}^{m}$ and $\mathbf{b}_{6}^{f}$ are vectors of fixed effects for male and female birds, respectively. Matrices of $\mathbf{X}_{1\text{-}5}^{m}$, $\mathbf{X}_{1\text{-}5}^{f}$, $\mathbf{Z}_{1\text{-}5}^{m}$, $\mathbf{Z}_{1\text{-}5}^{f}$, $\mathbf{W}_{1\text{-}5}^{m}$, $\mathbf{W}_{1\text{-}5}^{f}$, $\mathbf{X}_{6}^{m}$, $\mathbf{X}_{6}^{f}$, $\mathbf{Z}_{6}^{m}$ and $\mathbf{Z}_{6}^{f}$ are incidence matrices. Vectors $\mathbf{a}_{1\text{-}5}$ and $\mathbf{a}_{6}$ are direct additive genetic effects of the bird for BW1-2, WG3-6, respectively; $\mathbf{c}_{1\text{-}5}$ are vectors of the permanent environmental maternal effect for BW1-2 and WG3-5. Direct additive genetic and permanent environmental maternal effects were reduced ranks between sexes. Vectors $\mathbf{e}_{1\text{-}5}^{m}$, $\mathbf{e}_{1\text{-}5}^{f}$, $\mathbf{e}_{6}^{m}$ and $\mathbf{e}_{6}^{f}$ are random residuals of male BW1-2 and WG3-5, female BW1-2 and WG3-5, male WG6 and female WG6, respectively. The random vectors of the direct additive genetic effect and permanent environmental maternal effect were assumed to be normally distributed: $\boldsymbol{a}_{1-6}\sim \mathbf{N}[\mathbf{0}, \mathbf{A}\otimes \boldsymbol{V}_{a}^{0}]$, $\boldsymbol{c}_{1-5}\sim \mathbf{N}[\mathbf{0}, \mathbf{I}\otimes \boldsymbol{V}_{c}^{0}]$, where **A** is the pedigree relationship matrix; **I** is the identity matrix for dams; $\boldsymbol{V}_{a}^{0}$ is the 6x6 covariance matrix of the direct additive genetic effects; $\boldsymbol{V}_{c}^{0}$ is the 5x5 covariance matrix of the permanent environmental maternal effects. The residuals were also assumed to be normally distributed: $\mathbf{e}_{1\text{-}5}^{m}$, $\mathbf{e}_{1\text{-}5}^{f}$, $\mathbf{e}_{6}^{m}$ and $\mathbf{e}_{6}^{f}\sim \mathbf{N}[\mathbf{0}, \mathbf{I}\otimes \boldsymbol{V}_{e}^{0}]$, where I is the identity matrix for individual birds; the covariance matrix $\boldsymbol{V}_{e}^{0}$ is an 12x12 matrix, in which the residual covariances between male and female records for BW1-2 and WG3-5 traits are zero. Covariance matrices $\boldsymbol{V}_{a}^{0}$, $\boldsymbol{V}_{c}^{0}$ and $\boldsymbol{V}_{e}^{0}$ are at standardized scale.

Transformation of parameters to body weight scale

Parameters estimated from the model (4.3) were in standardized scale of BW and weight gains. However, results of parameters were presented in BW at different weeks of age by re-transforming (co)variances of standardized BW1-2 and WG3-6 back to the scale of BW1-6. Transformations from standardized BW1-2 and WG3-6 to normal scale of male and female BW1-6 were carried out for (co)variance matrices and the asymptotic covariance matrix. The asymptotic covariance matrix, which is the inverse of the average of observed and expected information in the REML likelihood (Jensen, 1997) from DMUAI procedures

(Madsen and Jensen, 2013), was used to compute approximate standard errors for estimates using the approach by Fischer *et al.* (2004). The transformation formula can be found in Appendix 4.1.

4.3. Results

Table 4.1 shows the means, coefficient of variation, standard deviation, and minimum and maximum values for BW at different weeks of age from male and female broilers reared in a commercial environment. The results show that BW of the birds increased drastically with increasing weeks of age. The increase was more than 2.5 times from BW1 to BW2. Similarly, the standard deviation of BW increased quickly at the early ages of birds, but the rate of the increase was lower at the later ages. The CV remained relatively stable over different ages of birds. The results also showed that both the absolute and the relative differences in BW between males and females increased with age. The relative differences between male and female BW was 1.01, 1.03, 1.06, 1.10, 1.12 and 1.15 at week 1-6, respectively.

Table 4.1: Descriptive statistics for body weight (BW) records (unit in gram) of males and females from commercial broiler chicken at 1-6 weeks of age.

Week	Male						Female					
	n	μ	CV	SD	Min.	Max.	n	μ	CV	SD	Min.	Max.
1	8039	155	0.14	22	68	236	8388	154	0.14	21	72	224
2	8631	403	0.16	63	152	582	9010	393	0.16	61	150	580
3	8425	810	0.16	131	278	1226	8870	762	0.15	116	302	1102
4	8225	1243	0.18	218	350	1930	8689	1131	0.17	189	336	1732
5	7455	1735	0.17	302	480	2796	7922	1550	0.16	248	506	2274
6	3975	2231	0.16	364	658	3225	4217	1940	0.15	290	685	2752

Note: n is number of records; μ is mean; CV is coefficient of variation; SD is standard deviation.

Table 4.2 shows the estimated variance components and their relative weekly increase. All variances increased sharply with increasing weeks of age. The relative increase in variances was larger at early age than later. For example, the relative increase in direct additive genetic variance from BW1 to BW2 was 1028 and 976% for males and females, respectively, whereas the relative increase in the variance from BW5 to BW6 was 51 and 41% for males and females. In addition, the increase of the direct genetic and residual variances was steeper than the increase of the maternal variance. The relative increase in variance of the permanent environmental maternal effect was lower than variances of the direct additive genetic effect at week 1-6. There was no increase in the permanent environmental maternal variance from week 5-6 as the effect was not significant for the weight gain from week 5-6.

The direct additive genetic, permanent environmental maternal and phenotypic variances were higher for male than female BW at all weeks of age. The relative increases in variances were also higher for male BW. The difference in variances between male and female BW increased with increasing weeks of age.

Table 4.2: Estimated variance components and relative increase in variance for body weight (BW) of broiler chicken over 1-6 weeks. Variance components were direct additive genetic effect (*a*), permanent environmental maternal effect (*c*), and phenotypic effect for male ([effect].M) and female BW ([effect].F). Relative increase in variances at week *t* (2-6) is the difference between variances at week *t* and week *t-1* divided by the variance at week *t-1*.

Week	Variance						Relative increase in variance (%)					
	a.M	a.F	c.M	c.F	p.M	p.F	a.M	a.F	c.M	c.F	p.M	p.F
1	86	83	33	32	306	300						
2	970	893	234	215	3089	2830	1028	976	609	572	909	843
3	3538	2821	786	636	12435	9712	265	216	236	196	303	243
4	10724	7936	1599	1231	34219	24668	203	181	103	94	175	154
5	23904	16733	2591	1911	74742	50852	123	111	62	55	118	106
6	36122	23621	2591	1911	130640	82065	51	41	0	0	75	61

Table 4.3 shows additive genetic, permanent environmental maternal and phenotypic coefficients of variation for male and female BW at different weeks of age. The direct additive genetic coefficients of variation fluctuated around 0.060-0.089 for male BW between week 1-6 and 0.059-0.083 for female BW between week 1-6. The permanent environmental maternal coefficients of variation had a decreasing tendency for both male and female BW as weeks of age increased from 1 to 6. In contrast, the environmental residual and phenotypic coefficients of variation had increasing tendency for male and female BW. The difference between phenotypic coefficients of variation of male and female BW increased with increasing weeks of age.

Table 4.3: Direct additive genetic (*a*), permanent environmental maternal (*c*) and phenotypic (*p*) coefficients of variation (*CV*) for male and female BW of broiler chicken over 1-6 weeks of age

BW at week	Male			Female		
	CV_a	*CV_c*	*CV_p*	*CV_a*	*CV_c*	*CV_p*
1	0.060	0.037	0.113	0.059	0.037	0.113
2	0.077	0.038	0.138	0.076	0.037	0.135
3	0.073	0.035	0.138	0.070	0.033	0.129
4	0.083	0.032	0.149	0.079	0.031	0.139
5	0.089	0.029	0.158	0.083	0.028	0.146
6	0.085	0.023	0.162	0.079	0.023	0.148

Figure 4.1 shows estimates of heritability and ratio of the permanent environmental maternal variance to the total phenotypic variance for BW over different weeks of age. With increasing weeks of age, the heritability had increasing tendency although there were two drops at week 3 and 6 when compared with heritability of BW in the previous week. Heritability of BW at week 1 and 6 were lowest at 0.276-0.288. Heritability of BW at week 5 was highest at 0.320 and 0.329 for male and female BW, respectively. The difference in heritability between male and female BW was negligible at all weeks of age, but the difference tended to increase with increasing weeks of age.

For BW, the ratio of the permanent environmental maternal variance to the total phenotypic variance reduced gradually from week 1-6. The ratio was 0.108 for male and 0.106 for female of BW at week 1, and it declined to 0.020 for male and 0.023 for female of BW at week 6. The difference in the permanent environmental maternal effect between male and female BW was negligible at all weeks of age. The permanent environmental maternal effect for WG6 was not included in the multivariate model because inclusion of the effect in the model led to convergence problems and the effect was not significant. However, the permanent environmental maternal effect of BW6 still existed because BW6 was the sum of BW2 and WG3-6.

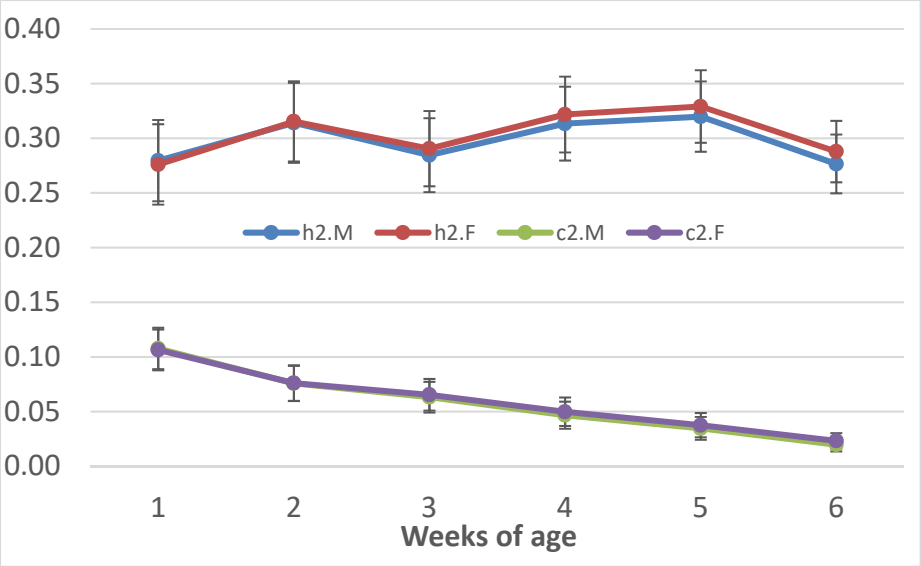


Figure 4.1: Heritability (h^2) (±SE) and ratio of the permanent environmental maternal variance to the total phenotypic variance (c^2) (±SE) for male (M) and female (F) BW over 1-6 weeks of age.

Table 4.4: Estimated genetic correlations of the direct additive genetic effect for male (above diagonal) and female BW (below diagonal) between BWs at different weeks of age.

Week	1	2	3	4	5	6
1	1	0.846	0.717	0.528	0.402	0.320
2	0.846	1	0.922	0.706	0.547	0.435
3	0.730	0.933	1	0.898	0.769	0.664
4	0.548	0.727	0.901	1	0.963	0.901
5	0.424	0.572	0.776	0.964	1	0.983
6	0.351	0.472	0.683	0.911	0.986	1

Note: Standard errors for the correlations ranged from 0.003-0.079.

Table 4.4 show the estimated genetic correlations of direct additive genetic effects between BWs at different weeks of age. Positive correlations between BWs at different weeks of age were found for the direct additive genetic effects. The direct additive genetic effects on two consecutive weekly BWs were highly correlated with genetic correlations ranging from 0.846-0.986 with standard error of 0.003-0.029. The correlations between two consecutive weekly BWs were lower in the early ages than in the late ages. The correlations between early and late BWs were weaker for the direct additive genetic effect. The genetic correlation between BW1 and BW6 was only 0.320 for male BW and 0.351 for female BW.

Table 4.5 show the phenotypic correlations between BWs at different weeks of age for male and female records. Similar to the genetic correlations between different weeks of age, the phenotypic correlations

were higher between two consecutive weekly BWs than the correlations between early and late BWs. The phenotypic correlations between two consecutive weekly BWs were 0.752-0.925 for male records and 0.763-0.941 for female records. The correlations between BWs at week 1-2 and week 5-6 were 0.307-0.546 for male records and 0.350-0.566 for female records. The correlations between BWs at different weeks of age for female BW were slightly higher than the corresponding correlations for male BW.

Table 4.5: Estimated phenotypic correlations between BWs at different weeks of age for males (above diagonal) and females (below diagonal)..

Week	1	2	3	4	5	6
1	1	0.752	0.635	0.502	0.395	0.307
2	0.763	1	0.869	0.688	0.546	0.431
3	0.664	0.885	1	0.863	0.724	0.595
4	0.521	0.697	0.864	1	0.915	0.797
5	0.421	0.566	0.732	0.921	1	0.925
6	0.350	0.468	0.625	0.831	0.941	1

Note: Standard errors for the correlations ranged from 0.002-0.013.

4.4. Discussion

A multivariate model was developed and used to estimate variance components for male and female BW at different weeks of age of broiler chicken tested in a commercial production environment. A criterion used in the development of the model was cross-validation procedure that was based on half-sib correlation in order to avoid biases due to maternal effects. The model used BW and weight gains as phenotypic records to overcome convergence problems, and male and female BW were standardized differently to model the heterogeneous variances between sexes.

Statistical model and methodology

A criterion for the development of statistical models was the predictive ability of breeding values in cross-validation tests that were based on the half-sib correlation (results not shown). Methods of estimating this predictability is different from the conventional method of cross-validation. The conventional method is based on correlation between corrected phenotypes (y_c) and EBVs of the same animals in the validation dataset (Christensen *et al.*, 2012). This conventional method can lead to an overestimation of the model predictive ability when maternal effects are present. The estimation of EBVs of birds in the validation dataset is from information of their full-sibs, half-sibs, dam, sire, and other relatives in the pedigree, of which full-sibs provide most information to the prediction. If the maternal effects are not accounted for

appropriately in the model, the effects shared among full-sibs may influence the EBVs of birds in the validation dataset, and the predictability of breeding values for the model would be overestimated.

To avoid this overestimation, we use a different approach to calculate the predictive ability in the cross validation including the division of training and validation datasets and half-sib correlation. The division is to ensure that birds in the validation dataset always have their full-sibs and half-sibs in the training dataset. If the maternal effects are not properly accounted for in the model, the effects will confound into EBVs. The correlation of EBV_v and y_c is between birds that are paternal half-sibs in the validation dataset, and due to the mating structure such half sibs will not share potential bias from maternal effects. However, the interference of the maternal additive genetic effect is reduced but not completely removed because the different dams of two half-sib birds might be related in the pedigree. Another limitation of our method is that the cross validation relies on correlation between EBV_v and y_c of half-sibs which may have higher standard error than the conventional method using correlation between EBVs and y_c of the same birds.

It is common that sex by genotype interaction is ignored in genetic evaluations of breeding programs because re-ranking due to sex by genotype interaction is small (van der Heide *et al.*, 2016), and because treating male and female traits as two traits can easily lead to convergence problems in the models. Heterogeneous variance between sexes, therefore, are often also ignored. However, when the heterogeneous variance exists and is not accounted in the model, a serious re-ranking may occur and lead to reduced response to selection (Cardoso *et al.*, 2007). Failure to account for different variances between sexes could also lead bias to variance components and estimated breeding values (Thompson, 2008). A distinct feature of the multivariate model (4.3) was the standardization applied differently to male and female records based on phenotypic standard deviations that were calculated from univariate models (1-2). This relatively simple standardization approach can model the heterogeneous variances between sexes and their unity correlations. This approach can reduce potential bias compared to the model that considered identical variances for sexes. In addition, the use of heterogeneous variances can be very good for multi-trait selection indexes for different sexes. Compared to the bivariate model that treats male and female records as two different traits, our model requires less computation and it is less likely to encounter convergence problems. Another feature of the multivariate model (4.3) was different residual variance for male and female traits. When the standardization of male and female records were not efficient, this would show in the model as a heterogeneous residual variance.

In addition, our model utilized all performances of BW1-6 simultaneously in the multivariate model. Because there were repeated measurement for BW, repeatability and random regression models were

considered. However, repeatability model would have low accuracy because correlations between early and late BW were low. Random regression models would require high order of fitting functions due to the fluctuation of heritability at different weeks of age, particularly at week 2-3. When convergence of the multivariate model was not an issue, the multivariate model had more parameters, and described better the covariance matrix of the direct additive genetic effect over weeks of age than the random regression model. In addition, the covariance matrices from a multivariate model can be fitted with functions, thus the change of an effect over week of ages can be described just like in a random regression model. A Legendre polynomial covariance function that were used to model the covariance of the additive genetic effect can be found in Appendix 4.2. Multivariate model is typically regarded as better prediction to breeding values than univariate model because the multi-trait model capitalize information from correlated traits (Henderson and Quaas, 1976).

Our model used BW1-2 and weight gains to model BW1-6. This is a linear transformation that leads to the same inferences but have much better convergence properties. However, after transformations, variance components of the model (4.3) were not as expected. Although correlations between sexes were assumed to be one, genetic correlations between BWs at different weeks were not identical for males and as females. For examples, correlation between BW1 and BW2 was 0.846 for both male and female, but correlation between male BW1 and male BW6 was different compared to the correlation between female BW1 and female BW6. Although the difference in heritability between sexes is negligible for BW3-6, there was an increasing difference in heritability between male and female with increasing weeks of age.

Variance components

With increasing age, the direct additive genetic, permanent environmental maternal and residual variances of BW increased sharply. However, the direct additive genetic, permanent environmental maternal and residual coefficients of variation changed relatively little with increasing weeks of ages. The small change in the coefficients of variation, despite of the sharp increase in variances, indicates that the change in variance components over week 1-6 is mainly due to scaling effect as the mean BW also increased considerably with increasing ages. The sharp increase in variances and relatively little changes in coefficients of variation were also found in BW of indigenous chicken from week 0-16 (Dana *et al.*, 2011), BW of crossbred chicken from week 2-10 (Begli *et al.*, 2016) and BW of broiler chicken from day *t-7* to *t* (Mebratie *et al.*, 2017).

However, the scaling effect was not the only factor responsible for the change of variance components over ages of broilers because the rates of the change were different between weeks of age and for different random effects in the model. The rate of the change in variance components from BW1 to BW2

is much more substantial than the change between later consecutive weeks of age. In addition, the rates of the change are different between different random effects. The permanent environmental maternal variance increased considerably but its proportion of the total phenotypic variance reduced gradually. The phenotypic variances also increased rapidly for both male and female BW, but the increase was at a slower rate for female BW due to the lower growth rate of females.

Maternal effects

The permanent environmental maternal effect on BW reduced gradually from week 1-6. The effect was still significant on BW at week 6, but the effect was not significant for WG6. Jasouri *et al.* (2017) also found a diminishing trend of the effect on BW in dual-purpose chicken when they aged. They found that the permanent environmental maternal effect was still significant at week 12. Dana *et al.* (2011) found the environmental maternal effect on BW at week 8, but not BW at week 12 in indigenous chicken. Begli *et al.* (2016) showed that the ratio of the permanent environmental maternal variance to the total phenotypic variance for BW of F2 chicken crossed between commercial broilers and native fowls increased slightly from 0.10 at week 2 to 0.12 at week 6, and then reduced to 0.07 at week 10. Maniatis *et al.* (2013) showed that the ratio of the permanent environmental maternal variance to the total phenotypic variance was 0.12 for BW at week 1 and 0.05 for BW at week 5 of commercial broiler chicken. The decrease in the permanent environmental maternal effect with increasing age was also found for meat quail (Barbieri *et al.*, 2015) and local Venda chicken (Norris and Ngambi, 2006). In comparisons, the permanent environmental maternal effect at the corresponding age in our study is lower than other studies (Begli *et al.*, 2016; Dana *et al.*, 2011; Jasouri *et al.*, 2017; Maniatis *et al.*, 2013; Zonuz *et al.*, 2013). This may be due to the raising environment. Birds in our study are raised in commercial production environment while those studies have birds from breeding units (Dana *et al.*, 2011; Jasouri *et al.*, 2017; Maniatis *et al.*, 2013; Zonuz *et al.*, 2013) or controlled experimental environments (Begli *et al.*, 2016; Norris and Ngambi, 2006).

The maternal effect, age of dam, was also included into the multivariate model (4.3), as it was in the study by Koerhuis and Thompson (1997) on juvenile BW of broilers. The influence of dam age on broiler BW may be related to egg weight. Significant effects of dam age and egg weight on broiler BW has been reported (Lapao *et al.*, 1999; Tahir *et al.*, 2011; Tona *et al.*, 2004; Wolanski *et al.*, 2007). The effect of egg weight was significant on BW of broilers at hatching and at 50 days of age, and a linear function of egg weight on hatching BW was found in Tahir *et al.* (2011). Tona *et al.* (2004) found that BWs of broilers at week 0-2 from younger dams were significantly lower than the BWs from older dams.

In our studies, the maternal additive genetic effect was not included in the multivariate model (4.3) as the effect was not significant for all BW traits except male BW1 (Appendix 4.3). Meanwhile, many studies show that the presence of the maternal effects including both additive genetic and permanent environmental effects can increase the fit of the models (Chapuis *et al.*, 1996; Jasouri *et al.*, 2017; Koerhuis and Thompson, 1997; Maniatis *et al.*, 2013; Zonuz *et al.*, 2013). Inclusion of both the maternal effects in the model improved considerably the fit of models for BW traits at week 0, 8 and 12 in dual-purpose chicken compared to the models without or with only one of the maternal effects (Jasouri *et al.*, 2017). A similar conclusion on the inclusion of both the maternal effects in the model was for BW at week 0 and 5 in broilers (Maniatis *et al.*, 2013), BW at week 8 in Iranian native chicken (Zonuz *et al.*, 2013), BW at week 12 and 16 in turkeys (Chapuis *et al.*, 1996), and juvenile BW in broilers (Koerhuis and Thompson, 1997). These five studies used REML for estimation of variance components, and comparison criteria between models was based on log-likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Jasouri *et al.* (2017) found that breeding values estimated for BW at week 0 and 8 were substantially affected by the maternal effects included in the model whereas breeding values estimated for BW at week 12 were similar for models with and without the maternal effects included.

Heritability

Heritability estimates in our study were moderate and ranged from 0.276 to 0.320 for male BW and from 0.276 to 0.329 for female BW at week 1-6. Kapell *et al.* (2012), in which broiler chicken were also raised in commercial production environment, obtained a heritability for BW5 of 0.32-0.36. In the same study (Kapell *et al.*, 2012), heritability of BW5 were 0.32-0.40 for broiler chicken that were raised in the breeding environment. Other studies on broiler chicken raised in a breeding environment report heritability of 0.21-0.42 for juvenile male BW, 0.30-0.53 for juvenile female BW (Mebratie *et al.*, 2017), 0.26-0.33 for juvenile male BW, 0.30-0.36 juvenile female BW (Koerhuis and Thompson, 1997) and 0.22 for BW at week 5 (Maniatis *et al.*, 2013). Also raised in the breeding environment, heritability were 0.19-0.23 for BW of Horro indigenous chicken at weeks 2-16 (Dana *et al.*, 2011), 0.31-0.32 for BW of dual-purpose chicken at weeks 8 and 12 (Jasouri *et al.*, 2017), and 0.24 for BW of Iranian native chicken at week 8 (Zonuz *et al.*, 2013). In Begli *et al.* (2016) crossbred chicken were raised in a controlled experimental environment, and heritability estimates varied from 0.32 to 0.34 for BW at week 2-6, and it dropped linearly from 0.34 for BW at week 6 to 0.19 for BW at week 10. In Norris and Ngambi (2006), heritability was 0.36 for BW of local Venda chicken at week 0 and 0.25 at week 4.

Heritability for BW at week 2-3 does not seem to follow the overall change of heritability over ages. It might be due to the change of maternal additive genetic effect at week 2-3 that is not accounted in the

multivariate model (4.3). Even if the maternal additive genetic effect was not significant, failing to account for this effect might cause an overestimation of the direct additive genetic effect. Other possible explanation to the change of heritability for BW at week 2-3 could be diseases. Birds around this age are prone to coccidiosis and other diseases of the digestive tract (Shamim *et al.*, 2015) that affect BW of broilers. Respiratory diseases caused by infectious bronchitis virus, avian metapneumovirus and mycoplasma species may also affect BW of broilers at this age, when level of maternal antibodies declines substantially and birds start to croak (De Boeck *et al.*, 2015). In addition, feed change from starter to growing diets may affect BW of broilers at this age.

Genetic correlations

The genetic correlation between BWs at consecutive weeks was high, particularly for BWs at weeks 2-6, in which the correlations between BWs at two consecutive weeks were above 0.898. The correlations between BWs at week 4 and 6 were also high at 0.901 for male records and 0.911 for female records. These findings are in agreement with Barbieri *et al.* (2015), Begli *et al.* (2016) and Niknafs *et al.* (2012), which estimated the genetic correlations between BWs at two consecutive weeks ranging from 0.90 to 0.99 for BWs from week 2 onwards. Begli *et al.* (2016) found high correlations (>0.91) between BWs at week 4-7. In Mebratie *et al.* (2017), the genetic correlation between juvenile BWs at *t* and *t+4* days of age were 0.94 for male records and 0.92 for female records. The genetic correlations were 0.97 between BWs at week 6 and 8, 0.86 between BWs at week 8 and 12, and 0.99 between BWs at week 12 and 16 in Dana *et al.* (2011). These high correlations imply that the age of selection for broilers can be reduced compared to the age of target BW in breeding program with relatively low loss in accuracy of selection. In our study, the correlations between early and late BWs were relatively low, ranging from 0.320 to 0.572 for BWs between week 1-2 and 5-6. Maniatis *et al.* (2013) found a genetic correlation of 0.17 between BWs at week 1 and 5. The genetic correlations for BWs were 0.66 between week 2 and 6, 0.46 between week 2 and 8, and 0.26 between week 2 and 10 (Begli *et al.*, 2016). Other studies found genetic correlations of BW to be 0.36-0.37 between week 1 and 8-12 (Niknafs *et al.*, 2012), 0.25-0.46 between week 0 and 6-16 (Dana *et al.*, 2011), and 0.20-0.38 between week 1 and 5-6 (Barbieri *et al.*, 2015). These findings suggest that early BW has a relatively low influence on target BW of selection.

Sex by genotype interaction

Sex by genotype interaction refers to the same genotype expressed differently in male and female BWs. This can be modelled by regarding BW in males and females as two different traits. Each bird, either male or female, would have two breeding values for male BW and for female BW. Indications of sex by genotype interaction can be heterogeneous variances, different heritability and correlation of less than one

between male and female BWs. Sex by genotype interaction for BW has been demonstrated in broilers (Mebratie *et al.*, 2017; van der Heide *et al.*, 2016) and other poultry species (Chapuis *et al.*, 1996; Mignon-Grasteau *et al.*, 1998). These studies show that variances between sexes could differ by a factor of 2 or more. These studies also found different heritability between sexes and correlations of 0.83-0.94 between male and female BW. However, these high correlations imply relatively small re-ranking between sexes. In addition, van der Heide *et al.* (2016) shows that correlations between estimated breeding values from sex-joined (univariate) and sex-specific (bivariate) models were 0.94-0.97 for broiler chicken BW at week 6.

In our study, the genetic variances for male and female BWs tended to be different, but the genetic correlation was not significantly different from one. Treating male and female BWs as two different traits led to convergence problems in the model, due to parameters at the edge of the parameter space. The unity correlation between sexes implies no re-ranking between male and female performances. The difference in heritability between sexes was also negligible even when different residuals for sexes were assumed in the multivariate model (4.3). Despite of unity correlation and similar heritability, the difference in genetic variances between male and female BW increased with increasing weeks of age.

4.5. Conclusion

A model was developed and used to estimate genetic parameters of BW at 1-6 weeks of age of broilers raised in a commercial environment. To improve accuracy of predicting EBVs, we have used several different approaches including model cross-validation based on half-sib correlation, scaling applied separately to male and female records and the use of weight gains to model BW. Half-sib correlation was used to reduce the interference of maternal effects on the cross-validation when maternal effects might be present. Scaling was to account for heterogeneous variance between sexes to reduce potential bias of the model that considered identical variances for sexes. All performances of BW1-6 were utilized simultaneously in a multivariate model using weight gains. Parameter estimates from the multivariate model show that the direct additive genetic, permanent environmental maternal and residual variances for BW increased sharply as age of broilers increased. The sharp increase in variances over weeks of age were mainly due to scaling effect. However, rate of the increase was also different e.g. ratio of the permanent environmental maternal variance to phenotypic variance reduced gradually with increasing age.

Appendix 4.1:

Transformations from standardized BW1-2 and WG3-6 to normal scale of male and female BW1-6 were carried out for (co)variance matrices and the asymptotic covariance matrices using formula (4.4-4.7):

$$\mathbf{V_a} = \mathbf{T_3}\left[\mathbf{T_2}\left(\mathbf{T_1}\,\mathbf{V_a^0}\,\mathbf{T_1'}\right)\mathbf{T_2'}\right]\mathbf{T_3'} \quad (4.4)$$

$$\mathbf{V_c} = \mathbf{T_{c3}}\left[\mathbf{T_{c2}}\left(\mathbf{T_{c1}}\,\mathbf{V_c^0}\,\mathbf{T_{c1}'}\right)\mathbf{T_{c2}'}\right]\mathbf{T_{c3}'} \quad (4.5)$$

$$\mathbf{V_e} = \mathbf{T_3}\left[\mathbf{T_2}\left(\mathbf{T_1}\,\mathbf{V_e^0}\,\mathbf{T_1'}\right)\mathbf{T_2'}\right]\mathbf{T_3'} \quad (4.6)$$

$$\mathbf{V_I} = \mathbf{T_{VI3}}\left[\mathbf{T_{VI2}}\left(\mathbf{T_{VI1}}\,\mathbf{V_I^0}\,\mathbf{T_{VI1}'}\right)\mathbf{T_{VI2}'}\right]\,\mathbf{T_{VI3}'} \quad (4.7)$$

where matrices of direct additive genetic, permanent environmental maternal, residual and asymptotic covariances were Va, Vc, Ve and VI, respectively, at normal BW scale, and $\mathbf{V_a^0}$, $\mathbf{V_c^0}$, $\mathbf{V_e^0}$ and $\mathbf{V_I^0}$, respectively, at standardized weight gain scale. Transforming matrices for formula (4-7) were:

$$\mathbf{T_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{T_3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{T_{c1}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{T_{c3}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{T}_{\mathrm{VI1}} = \begin{bmatrix} \mathbf{T_1} \otimes \mathbf{T_1} & 0 & 0 \\ 0 & \mathbf{T_{c1}} \otimes \mathbf{T_{c1}} & 0 \\ 0 & 0 & \mathbf{T_1} \otimes \mathbf{T_1} \end{bmatrix}$$

$$\mathbf{T}_{\mathrm{VI2}} = \begin{bmatrix} \mathbf{T_2} \otimes \mathbf{T_2} & 0 & 0 \\ 0 & \mathbf{T_{c2}} \otimes \mathbf{T_{c2}} & 0 \\ 0 & 0 & \mathbf{T_2} \otimes \mathbf{T_2} \end{bmatrix}, \mathbf{T}_{\mathrm{VI3}} = \begin{bmatrix} \mathbf{T_3} \otimes \mathbf{T_3} & 0 & 0 \\ 0 & \mathbf{T_{c3}} \otimes \mathbf{T_{c3}} & 0 \\ 0 & 0 & \mathbf{T_3} \otimes \mathbf{T_3} \end{bmatrix}$$

Matrix $\mathbf{T_2}$ is a 12x12 matrix, of which off-diagonal elements are zero, the diagonal is vector of phenotypic standard deviations with trait orders: male BW1, female BW1, male BW2, female BW2, male WG3, female WG3, male WG4, female WG4, male WG5, female WG5, male WG6 and female WG6. Matrix $\mathbf{T_{c2}}$ is a 10x10 matrix that is sub-matrix of $\mathbf{T_2}$ for traits of male and female BW1-2 and WG3-5. The phenotypic standard deviations of diagonals from matrices $\mathbf{T_2}$ and $\mathbf{T_{c2}}$ were computed from variance estimates of univariate models (4.1) and (4.2).

Appendix 4.2:

The covariance of a Legendre polynomial function that was fitted to model the additive genetic variances and covariances of BW traits over 1-6 week of ages was computed:

$$\widehat{\mathbf{V}_\mathbf{a}^\mathbf{L}} = \boldsymbol{\phi}^{-1}\widehat{\mathbf{V}_\mathbf{a}}\boldsymbol{\phi}^{-T} \qquad (4.8)$$

where $\widehat{\mathbf{V}_\mathbf{a}^\mathbf{L}}$ is the covariance matrix of the Legendre polynomial coefficients for the additive genetic effect of BW traits over weeks of age; $\boldsymbol{\phi}$ is a matrix of the Legendre polynomial coefficients with order of five that were computed using standardized weeks of age; $\boldsymbol{\phi}^{-1}$ is the inverse of $\boldsymbol{\phi}$; $\boldsymbol{\phi}^{-T}$ is the transpose of the inverse of $\boldsymbol{\phi}$; and $\widehat{\mathbf{V}_\mathbf{a}}$ is the covariance matrix of the additive genetic effect for BW1-6 estimated from the multivariate model (4.3).

Variances and covariances of the Legendre polynomial coefficients on weeks were computed separately for male and female BW (Table 4.6).

Table 4.6: Variances and covariances of Legendre polynomial coefficients on weeks for additive genetic effects of male and female body weights

	Male						Female					
	a_0	a_1	a_2	a_3	a_4	a_5	a_0	a_1	a_2	a_3	a_4	a_5
a_0	14143						10369					
a_1	9677	7630					6664	5063				
a_2	945	1347	639				464	790	394			
a_3	-1008	-815	-112	145			-726	-595	-87	109		
a_4	-284	-286	-106	26	25		-215	-207	-71	21	18	
a_5	147	140	29	-42	-5	21	124	110	19	-32	-4	15

Notes: a_0 is a coefficient on intercept; a_1, a_2, a_3, a_4 and a_5 are a coefficient on weeks to the first, second, third, fourth and fifth power, respectively.

Appendix 4.3:

Log-likelihood ratio tests (significant difference, P<0.05) were carried out to identify the significance of maternal additive genetic effect for BW traits by week and sex. Three univariate models (4.9-4.11) were used for the log-likelihood ratio tests:

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Za}+\mathbf{Mm}+\mathbf{Wc}+\mathbf{e} \quad (4.9)$$

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Za}+\mathbf{Mm}+\mathbf{Wc}+\mathbf{e} \quad (4.10)$$

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Za}+\mathbf{Wc}+\mathbf{e} \quad (4.11)$$

where **y** is vector of male or female phenotypic records of BW1-6 at normal scale; **b** is vectors of fixed effects of flock of birds, source of flocks and age of dam. Matrices of **X**, **Z**, **M**, and **W** are incidence matrices. Vectors **a**, **m**, **c** and **e** are the direct additive genetic effect, maternal additive genetic effect, permanent environmental maternal effect and residual, respectively. In model (4.9), these random effects were assumed to be normally distributed: $\begin{bmatrix}\mathbf{a}\\ \mathbf{m}\end{bmatrix} \sim \mathbf{N}[\mathbf{0}, \mathbf{A} \otimes \begin{bmatrix}\sigma_a^2 & \sigma_{a,m}\\ \sigma_{a,m} & \sigma_m^2\end{bmatrix}$, $\mathbf{c} \sim \mathbf{N}[\mathbf{0}, \mathbf{I_d}\sigma_c^2]$ and $\mathbf{e} \sim \mathbf{N}[\mathbf{0}, \mathbf{I}\sigma_e^2]$, where **A** is the pedigree relationship matrix; $\mathbf{I_d}$ is the identity matrix for dams; **I** is the identity matrix for individual birds; σ_a^2, σ_m^2, $\sigma_{a,m}$, σ_c^2 and σ_e^2 are the direct additive genetic variance, maternal additive genetic variance, direct and maternal additive genetic covariance, permanent environmental maternal variance, and residual variance, respectively. The random effects in model (4.10) were assumed to be normally distributed as in model (4.9), except that $\begin{bmatrix}\mathbf{a}\\ \mathbf{m}\end{bmatrix} \sim \mathbf{N}[\mathbf{0}, \mathbf{A} \otimes \begin{bmatrix}\sigma_a^2 & 0\\ 0 & \sigma_m^2\end{bmatrix}$. The random effects in

model (4.11) were assumed to be normally distributed as in model (9), except that the maternal additive genetic effect is not present.

Table 4.7: Estimates (±SE) of variance components[1] from model (4.9) and the significance of maternal additive genetic effects from log-likelihood ratio tests for male (M) and female (F) body weight of broiler chicken over 1-6 weeks of age.

Week	Sex	σ_a^2	$\sigma_{a,m}$	σ_m^2	σ_c^2	σ_e^2	Significance[2] of $\sigma_{a,m}$	Significance[3] of σ_m^2
1	M	64 ±14	-6 ±10	28 ±10	18 ±6	201 ±8	NS	***
1	F	64 ±14	0 ±8	12 ±7	24 ±5	198 ±8	NS	NS
2	M	816 ±154	13 ±80	65 ±58	208 ±48	1905 ±86	NS	NS
2	F	789 ±141	39 ±66	46 ±47	155 ±43	1737 ±79	NS	NS
3	M	3035 ±578	-53 ±298	225 ±224	773 ±186	7707 ±329	NS	NS
3	F	2890 ±491	-83 ±231	181 ±165	395 ±133	5800 ±274	NS	NS
4	M	10622 ±1793	-713 ±872	695 ±618	1585 ±472	19284 ±993	NS	NS
4	F	7786 ±1279	-220 ±596	442 ±414	892 ±339	14147 ±709	NS	NS
5	M	20665 ±3567	52 ±1628	836 ±1077	1986 ±978	44490 ±2042	NS	NS
5	F	15786 ±2654	358 ±1207	536 ±782	1817 ±746	29082 ±1484	NS	NS
6[4]	M	25558 ±6187		1014 ±2510	2530 ±2399	87633 ±3968	NS	NS
6[4]	F	23450 ±4667		0 ±1606	802 ±1413	46161 ±2718	NS	NS

Notes:

[1] Variance components estimated from model (4.9) were direct additive genetic variance (σ_a^2), maternal additive genetic variance (σ_m^2), covariance between direct and maternal additive genetic effects ($\sigma_{a,m}$), permanent environmental maternal variance (σ_c^2), and residual variance (σ_e^2).

[2] Log-likelihood ratio tests (significant difference, P<0.05) were used to test the significance of $\sigma_{a,m}$ by comparing log-likelihoods of models (4.9) and (4.10).

[3] Log-likelihood ratio tests (significant difference, P<0.05) were used to test the significance of σ_m^2 by comparing log-likelihoods of models (4.9) and (4.11).

NS: no significant difference (P >0.05); ***: significant difference with P<0.001.

[4] Because model (4.9) did not converge, variance estimates for BW6 were from model (4.10). There was no estimate for $\sigma_{a,m}$ in model (4.10).

References:

Barbieri, A, Ono, RK, Cursino, LL, Farah, MM, Pires, MP, Bertipaglia, TS, Pires, AV, Cavani, L, Carreno, LO & Fonseca, R 2015, 'Genetic parameters for body weight in meat quail', *Poult Sci,* vol. 94, no. 2, pp. 169-71.

Begli, HE, Torshizi, RV, Masoudi, AA, Ehsani, A & Jensen, J 2016, 'Longitudinal analysis of body weight, feed intake and residual feed intake in F_2 chickens', *Livestock Science,* vol. 184, no. 28-34.

Besbes, B & Ducrocq, V 2003, 'Use of mixed model methodology in breeding strategies for layers'. *In* '*Poultry Genetics, Breeding and Biotechnology*' (W. M. Muir and S. E. Aggrey, eds.), CABI, pp. 127-146.

Cardoso, FF, Rosa, GJ & Tempelman, RJ 2007, 'Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of postweaning gain of Nelore-Hereford cattle', *J Anim Sci,* vol. 85, no. 4, pp. 909-18.

Chapuis, H, Tixier-Boichard, M, Delabrosse, Y & Ducrocq, V 1996, 'Multivariate restricted maximum likelihood estimation of genetic parameters for production traits in three selected turkey strains', *Genetics Selection Evolution,* vol. 28, no. 3, pp. 299.

Christensen, OF, Madsen, P, Nielsen, B, Ostersen, T & Su, G 2012, 'Single-step methods for genomic evaluation in pigs', *Animal,* vol. 6, no. 10, pp. 1565-71.

Dana, N, vander Waaij, EH & van Arendonk, JAM 2011, 'Genetic and phenotypic parameter estimates for body weights and egg production in Horro chicken of Ethiopia', *Tropical Animal Health and Production,* vol. 43, no. 1, pp. 21-28.

De Boeck, C, Kalmar, I, Dumont, A & Vanrompay, D 2015, 'Longitudinal monitoring for respiratory pathogens in broiler chickens reveals co-infection of Chlamydia psittaci and Ornithobacterium rhinotracheale', *J Med Microbiol,* vol. 64, no. Pt 5, pp. 565-74.

Di Masso, RJ, Dottavio, AM, Canet, ZE & Font, MT 1998, 'Body weight and egg weight dynamics in layers', *Poult Sci,* vol. 77, no. 6, pp. 791-6.

Fischer, TM, Gilmour, AR & van der Werf, JHJ 2004, 'Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML', *Genetics Selection Evolution,* vol. 36, no. 3, pp. 363.

Grosso, JL, Balieiro, JC, Eler, JP, Ferraz, JB, Mattos, EC & Michelan Filho, T 2010, 'Comparison of different models to estimate genetic parameters for carcass traits in a commercial broiler line', *Genet Mol Res,* vol. 9, no. 2, pp. 908-18.

Henderson, C & Quaas, R 1976, 'Multiple trait evaluation using relatives' records', *Journal of Animal Science,* vol. 43, no. 6, pp. 1188-1197.

Jasouri, M, Zamani, P & Alijani, S 2017, 'Dominance genetic and maternal effects for genetic evaluation of egg production traits in dual-purpose chickens', *Br Poult Sci,* vol. 58, no. 5, pp. 498-505.

Jensen, J 1997, 'Residual maximum likelihood estimation of (co) variance components in multivariate mixed linear models using average information', *Journal of the Indian Society of Agricultural Statistics,* vol. 49, no. 215-236.

Kapell, DN, Hill, WG, Neeteson, AM, McAdam, J, Koerhuis, AN & Avendano, S 2012, 'Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments', *Poultry Science,* vol. 91, no. 3, pp. 565-74.

Koerhuis, AN & Thompson, R 1997, 'Models to estimate maternal effects for juvenile bodyweight in broiler chickens', *Genet Sel Evol,* vol. 29, no.

Lapao, C, Gama, LT & Soares, MC 1999, 'Effects of broiler breeder age and length of egg storage on albumen characteristics and hatchability', *Poultry Science,* vol. 78, no. 5, pp. 640-645.

Madsen, P & Jensen, J 2013, 'DMU: a user's guide'. "*A package for analysing multivariate mixed models, Version 6, release 5.2. Home page: http://dmu.agrsci.dk/*",

Maniatis, G, Demiris, N, Kranis, A, Banos, G & Kominakis, A 2013, 'Model comparison and estimation of genetic parameters for body weight in commercial broilers', *Canadian journal of animal science,* vol. 93, no. 1, pp. 67-77.

Mebratie, W, Shirali, M, Madsen, P, Sapp, R, Hawken, R & Jensen, J 2017, 'The effect of selection and sex on genetic parameters of body weight at different ages in a commercial broiler chicken population', *Livestock Science,* vol. 204, no. 78-87.

Mignon-Grasteau, S, Beaumont, C, Poivey, J-P & De Rochambeau, H 1998, 'Estimation of the genetic parameters of sexual dimorphism of body weight in'label'chickens and Muscovy ducks', *Genetics Selection Evolution,* vol. 30, no. 5, pp. 481.

N'Dri, AL, Sellier, N, Tixier-Boichard, M, Beaumont, C & Mignon-Grasteau, S 2007, 'Genotype by environment interactions in relation to growth traits in slow growing chickens', *Genetics Selection Evolution,* vol. 39, no. 5, pp. 513-528.

Niknafs, S, Nejati-Javaremi, A, Mehrabani-Yeganeh, H & Fatemi, SA 2012, 'Estimation of genetic parameters for body weight and egg production traits in Mazandaran native chicken', *Tropical Animal Health and Production,* vol. 44, no. 7, pp. 1437-1443.

Norris, D & Ngambi, JW 2006, 'Genetic parameter estimates for body weight in local Venda chickens', *Trop Anim Health Prod,* vol. 38, no. 7-8, pp. 605-9.

Roehe, R & Kennedy, B 1993, 'The influence of maternal effects on accuracy of evaluation of litter size in swine', *Journal of animal science,* vol. 71, no. 9, pp. 2353-2364.

Rovadoscki, GA, Petrini, J, Ramirez-Diaz, J, Pertile, SF, Pertille, F, Salvian, M, Iung, LH, Rodriguez, MAP, Zampar, A & Gaya, LG 2016, 'Genetic parameters for growth characteristics of free-range chickens under univariate random regression models', *Poultry science,* vol. 95, no. 9, pp. 1989-1998.

Shamim, A, ul Hassan, M, Yousaf, A, Iqbal, MF, Zafar, MA, Siddique, RM & Abubakar, M 2015, 'Occurrence and identification of Emeria species in broiler rearing under traditional system', *Journal of Animal Science and Technology,* vol. 57, no. 41.

Tahir, M, Cervantes, H, Farmer, CW, Shim, MY & Pesti, GM 2011, 'Broiler performance, hatching egg, and age relationships of progeny from standard and dwarf broiler dams', *Poultry Science,* vol. 90, no. 6, pp. 1364-1370.

Thompson, R 2008, 'Estimation of quantitative genetic parameters', *Proceedings of the Royal Society of London B: Biological Sciences,* vol. 275, no. 1635, pp. 679-686.

Tona, K, Onagbesan, O, De Ketelaere, B, Decuypere, E & Bruggeman, V 2004, 'Effects of age of broiler breeders and egg storage on egg quality, hatchability, chick quality, chick weight, and chick posthatch growth to forty-two days', *Journal of Applied Poultry Research,* vol. 13, no. 1, pp. 10-18.

van der Heide, E, Lourenco, D, Chen, C-Y, O. Herring, W, L. Sapp, R, Moser, D, Tsuruta, S, Masuda, Y, Ducro, B & Misztal, I 2016, 'Sexual dimorphism in livestock species selected for economically important traits', *Journal of Animal science,* vol. 94, no. 9, pp. 3684-3692.

Whiting, T & Pesti, G 1984, 'Broiler performance and hatching egg weight to marketing weight relationships of progeny from standard and dwarf broiler dams', *Poultry Science,* vol. 63, no. 3, pp. 425-429.

Wolanski, NJ, Renema, RA, Robinson, FE, Carney, VL & Fancher, BI 2007, 'Relationships Among Egg Characteristics, Chick Measurements, and Early Growth Traits in Ten Broiler Breeder Strains', *Poultry Science,* vol. 86, no. 8, pp. 1784-1792.

Zonuz, YA, Alijani, S, Rafat, S, Abbasi, M & Kia, HD 2013, 'Estimation of maternal effects on the North-Iranian native chicken traits using Bayesian and REML methods', *Slovak J. Anim. Sci.,* vol. 46, no. 52-60.

5

On farm observations to increase genetic gain in breeding schemes for village poultry production – A simulation study

Thinh Tuan Chu[1,2,4], John W.M. Bastiaansen[2], Peer Berg[1,3] & Elise Norberg[1,3]

[1] *Aarhus University, 8830 Tjele, Denmark*

[2] *Wageningen University & Research Animal Breeding and Genomics, 6709 PG Wageningen, The Netherlands*

[3] *Norwegian University of Life Sciences, 1433 Ås, Norway*

[4] *Corresponding author's email: chu.thinh@mbg.au.dk*

Published in *Acta Agriculturae Scandinavica, Section A — Animal Science*, 2018, 68(1):1:10

https://doi.org/10.1080/09064702.2018.1543444

Abstract

To improve genetic gain of breeding programs for village poultry production, breeding schemes with observations obtained in village production systems using individual (VIO) and group recording (VGO) were examined under different levels of genotype-by-environment-interactions (GxE). GxE was modelled by varying the correlation between traits measured in the breeding station and village environments for bodyweight (r_{g_BW}) and egg production (r_{g_EP}). Relative and absolute genetic gains obtained from VIO and VGO were used for comparison between the schemes. Results showed that village observations significantly improved genetic gains compared to the scheme without birds tested in the village. The improvement was only slightly larger with individual observations than with group observations. Higher r_{g_BW} and r_{g_EP} led to lower relative genetic gain, but higher absolute gain of VIO and VGO. It is recommended to apply a breeding scheme using group recording of village performance when strong GxE in breeding for village poultry is expected.

Key words: breeding scheme; GxE; group recording; stochastic simulation; village poultry

5.1. Introduction

Introduction of exotic breeds or high yielding hybrids has failed to upgrade the genetic level of the current chicken populations in Ethiopia due to various reasons, such as farmer preference, lack of required input and chicken adaptability (Dana et al., 2010; Wondmeneh et al., 2015). In addition, the application of exotic breeds in an intensive or semi-intensive production system for smallholder villagers brings in lower economic returns than the use of indigenous chicken under a scavenging production system (FAO, 2010; Okeno et al., 2013). Moreover, one of the biggest advantages of indigenous chicken is their disease resistance and adaptability to harsh conditions (Dessie et al., 2000). Therefore, a key approach for delivering a productive and adapted chicken suitable for the production system and acceptable to the farmers, is to improve the indigenous chicken through breeding programs.

A selective breeding program was initiated in 2008 at the Debre Zeit Agricultural Research Centre in Ethiopia (Dana et al., 2011). The ultimate objective of the breeding program is an improved dual-purpose chicken (Horro) for growth and egg production, which also is well-adapted to the semi-scavenging environment of village poultry production. However, the breeding scheme of the program has revealed to be suboptimal as it has shown slow genetic progress and signs of losing adaptability of indigenous chicken after 7 generations of selection (Wondmeneh, 2015). The differences between the conditions at the research station and villages might cause genotype by environment interaction (GxE). At the research

station, birds are selected under hygienic conditions, nutritionally adequate diets and well-protected cages, whereas at the villages, birds are subjected to a combination of low food availability, sub-optimal diet, prevalence of diseases and other social interaction factors.

Significant GxE in poultry has been reported in a number of studies (Bekele et al., 2009; Chen et al., 2009; Horst, 1985; Kapell et al., 2012; Mathur & Horst, 1994; N'Dri et al., 2007) and reviews (FAO, 2010; Mathur, 2003). GxE could reduce potential genetic gains of a breeding program. There are, however, only a few studies (Bijma & Arendonk, 1998; Mulder & Bijma, 2005) on design or evaluation of breeding schemes in the presence of GxE, and they are mainly designed for other species than poultry and for commercial production instead of village production. A big challenge for implementing breeding schemes for village poultry is the need for routine collection of observations on individual animals. Group mean of full-sibs and half-sibs can be a possible alternative for village phenotype recording. Studies on pooled data has illustrated that selection based on estimated breeding values (EBV) from pooled observations can be effective, particularly when group members have close relationships (Biscarini et al., 2008; Nurgiartiningsih et al., 2004; Olson et al., 2006; Peeters et al., 2013). However, the use of pooled observations in breeding programs where GxE is present, and where animals with the pooled observations are not candidates of selection, has not been demonstrated.

This paper proposes breeding schemes for village dual-purpose poultry production in the presence of GxE. Stochastic simulation is applied to compare breeding schemes on genetic gain considering group and individual recording and to optimize the data recording effort in villages versus stations. GxE was modelled by varying the correlation between traits measured in station and village environments.

5.2. Materials and methods

Breeding schemes

The stochastic simulation program ADAM (Pedersen et al., 2009) was used to simulate 100 replicates for each scenario. The simulation mimicked the situation of the Horro chicken breeding population at the Debre Zeit Agricultural Research Centre, Ethiopia (Dana et al., 2011). The schemes were designed for dual-purpose village poultry production, by including body weight (BW) and egg production (EP) in the breeding goal (Figure 5.1). The breeding structure consisted of 30 roosters and 300 hens. In each generation, a hen had 4 offspring that were candidates for selection and an additional number of offspring for testing. Sex was randomly assigned to offspring at a 50:50 ratio. The candidates for selection were reared in a research station. Under the station conditions, the birds had phenotypes defined as “station” traits. Birds for testing were transferred to village small holders for recording of phenotypes, which were defined as “village”

traits. The village tested birds were not considered as selection candidates, but only gave information for evaluating station selection candidates.

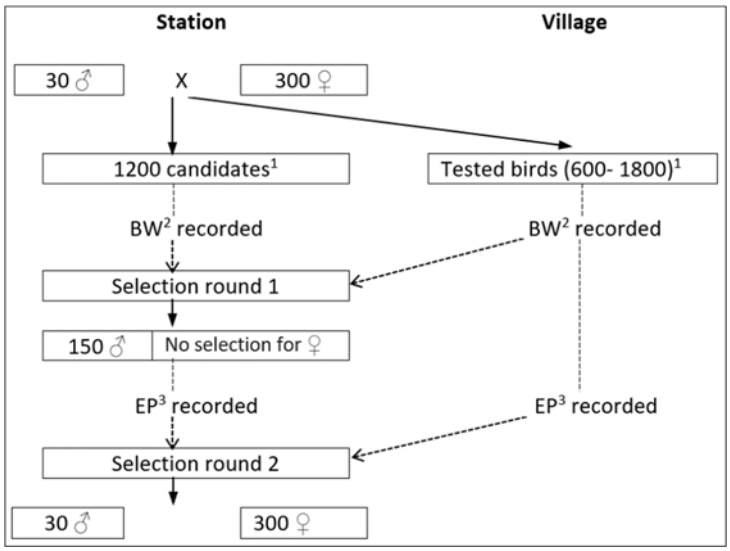


Figure 5.1: Breeding cycle of a generation. [1] Sex ratio of 1♂:1♀; [2] Bodyweight observed in both ♂ and ♀; [3] Egg production observed in ♀; → Birds reproduced/ selected, --- Observations realized, ⇢

Information for selection

In each generation, selection candidates went through two selection rounds. In the first selection round, 150 of all male candidates were selected after phenotypes for BW were measured both in station and in village. No selection was applied in the females. This round was to ensure a high selection response and to reduce costs of keeping all male candidates until EP was recorded. In the second round, 30 males were selected out of the remaining 150 candidates and 300 females were selected out of all female candidates. Selection round 2 was performed after phenotypes for EP were realized. BLUP selection was applied for both the selection rounds, so information about relatives both in station and in village was used. Selection was simulated for 20 discrete generations.

Trait simulation

Phenotypes of BW and EP were simulated and BW was observed for both male and female birds while EP was observed in females only. In the station environment, BW and EP were denoted as BW_s and EP_s, respectively, whereas in the village environment, the phenotypes were denoted as BW_v and EP_v, respectively. Observations on BW_s and EP_s were realized individually, while BW_v and EP_v were recorded as either group mean or individually. Group records were the average of the simulated phenotypes of 10 paternal-sibs, which were randomly selected from the 40 offspring of a sire. Therefore, members of a group could have both full-sib and half-sib relationship.

The genetic parameters assumed for all traits are shown in Table 5.1. The parameters of (co)variances, correlation and heritabilities of BW_s and EP_s were based on literature for indigenous chicken in Africa (Dana et al., 2011; Lwelamira et al., 2009; Niknafs et al., 2012; Oleforuh-Okoleh, 2011). We assumed that the village scavenging system would result in a larger environmental variance and a lower heritability compared to the conditions on station. The heritabilities for village traits was set to half the values for the station traits. Additive genetic variances of BW_s and EP_s were assumed equal to those of BW_v and EP_v, respectively. Genetic correlation between BW_s and EP_s was also equal to that of BW_v and EP_v. Genetic correlations between the village and station environments for BW (r_{g_BW}) and EP (r_{g_EP}) were varied to reflect different extent of GxE. To ensure a positive-definite matrix of genetic covariance, the genetic correlation between BW_s and EP_v was approximated by multiplying the average of r_{g_BW} and r_{g_EP} by correlation between BW_s and EP_s (Table 5.1). This approximation came from assuming that the link between BW_s and EP_v might be through either one of two paths. One path was through correlation between BW_s and EP_s and correlation between EP_s and EP_v, and another path was through correlation between BW_s and BW_v and correlation between BW_v and EP_v. The genetic correlation between BW_v and EP_s was approximated in the same way. Environmental correlation between BW_v and EP_v was assumed to be equal to that between BW_s and EP_s. Other environmental correlations between traits were set to 0 because birds only had records either on the station or in the village environment.

Table 5.1: Genetic parameters assumed for simulating body weight (BW) and egg number (EP) in station (s) and village (v) environments: phenotypic variance, heritability (along the diagonal), genetic correlations (above diagonal), and environmental correlations (below diagonal)

	σ_P^2	BW_s	EP_s	BW_v	EP_v
BW_s	291751	0.41	-0.12	r_{g_BW}	-0.12 (r_{g_BW} + r_{g_EP})/2
EP_s	130.65	0.02	0.28	-0.12 (r_{g_BW} + r_{g_EP})/2	r_{g_EP}
BW_v	569610	0	0	0.21	-0.12
EP_v	261.29	0	0	0.02	0.14

Note: r_{g_BW} and r_{g_EP}, genetic correlation between traits of station and village environments, are variable factors.

True breeding values of BW_s, EP_s, BW_v and EP_v traits of a bird *i* at generation 0 were scaled to achieve an initial genetic covariance matrix by following equation: $\mathbf{tbv_i} = \mathbf{L}' \times \mathbf{r}$, where $\mathbf{tbv_i}$ is a vector of true breeding values of bird *i*; $\mathbf{L}'$ is the Cholesky decomposition of the initial genetic covariance matrix; and **r** is a vector of random numbers from a standardized normal distribution. Means of the traits were 0. Simulation of environmental values of the traits was similar to simulation of true breeding values, with a

Cholesky decomposition of the environmental covariance matrix. Phenotypic observation of a trait for an individual was the sum of true breeding value and environmental value. Environmental (co)variances were kept constant through the simulations whereas genetic (co)variance and heritability decreased due to Bulmer effect of selection and inbreeding. True breeding value of the descendants was half of true breeding values of their parents plus Mendelian sampling terms. Mendelian sampling variance of the offspring was determined based on the inbreeding of the parents.

Simulation of group mean observations was done in two steps. The first step was simulation of individual phenotypic observations as described above. The second step was to compute group mean observations. All offspring birds of a sire in a village were randomly assigned into groups of 10 birds. Individual phenotypic observations of those offspring birds were used to calculate group means. Subsequently, the individual phenotypic observations were replaced by group means. For BW_v, 10 paternal-sib of a group had the same group mean observation. For EP_v, phenotypic observations of females of the 10 paternal-sib group were used to calculate the group mean, and phenotypic observations of these females were replaced by the mean.

Simulated scenarios

A reference breeding scheme and 2 alternative breeding schemes were simulated (Table 5.2). The reference breeding scheme had 1200 candidates for selection and 600, 1200 or 1800 tested birds. Both the selection candidates and tested birds provided information of station phenotypes. For the two alternative schemes, the tested birds did not provide information of station phenotypes but were transferred to village environment to get village phenotypes. In one of the alternatives (breeding scheme VIO), the village birds had individual observations. In the other (breeding scheme VGO), the birds had group mean observation of 10 paternal-sibs. As suggested in Cahaner et al. (1993), Kapell et al. (2012), Mathur and Horst (1994), Mathur (2003) and Chen et al. (2009), a stronger GxE interaction was simulated for EP than for BW, and therefore a lower genetic correlation between station and village measures. The lower correlation for EP than BW came from assumption that traits of reproduction have stronger GxE interaction than traits of production, and traits with lower heritability generally display higher GxE (Mathur, 2003). The values of r_{g_BW} were set at 0.5, 0.7 and 0.9 and r_{g_EP} were 0.1, 0.3 and 0.5.

Table 5.2: Breeding schemes and parameters of genetic correlations between station and village traits

Variables	Reference breeding scheme	Alternative breeding scheme	
		Individual observation (VIO)	Group observation (VGO)
Number of tested birds	600, 1200, 1800	600, 1200, 1800	600, 1200, 1800
Type of observations on tested birds	Station	Village	Village
Recording method	Individual	Individual	Group
Genetic correlation between station and village bodyweight (r_{g_BW})	0.5, 0.7, 0.9	0.5, 0.7, 0.9	0.5, 0.7, 0.9
Genetic correlation between station and village egg production (r_{g_EP})	0.1, 0.3, 0.5	0.1, 0.3, 0.5	0.1, 0.3, 0.5

As a consequence, there were 4 factors investigated: type of breeding schemes, number of tested birds, r_{g_BW} and r_{g_EP}. All three breeding schemes were simulated with all three numbers of tested birds and all 9 combinations of r_{g_BW} and r_{g_EP} resulting in a total of 81 simulated scenarios.

Selection criteria

Breeding was done to optimize production in the village environment and therefore the breeding goal was as follows:

$H = 0^* BW_s + 0^* EP_s + 0.078^* BW_v + 9.080^* EP_v$ (5.1)

An economic value of 0 was assigned to the station traits of BW_s and EP_s with the assumption that only village performance mattered. Economic values given to BW_v and EP_v were from Okeno et al. (2012). Unit of BW was measured in grammes, and EP was cumulative number of eggs produced until 40 weeks of age. Breeding values were estimated based on data from VIO and VGO using multivariate best linear unbiased prediction (BLUP) models. For individual phenotypic observation, the model was:

$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$ (5.2)

where **y** is a vector of individual phenotypic records of traits of BW_s, EP_s, BW_v and EP_v; **b** is a vector of fixed year effects; **a** is a vector of animal breeding values of the traits to be estimated assumed $\boldsymbol{a} \sim MVN[\mathbf{0}, \mathbf{A} \otimes \mathbf{G}]$, where *MVN* is the multivariate normal distribution, **A** is the additive genetic relationship matrix among individuals and **G** is the additive genetic (co)variance matrix among the traits as a 4x4 matrix; **X** and **Z** are incidence matrices relating fixed effects and breeding values to phenotypic observations of birds; and **e** is a vector of residuals of the traits assumed $\mathbf{e} \sim MVN\begin{bmatrix} 0 & \mathbf{I_s} \otimes \mathbf{E_s} & 0 \\ 0' & 0 & \mathbf{I_v} \otimes \mathbf{E_v} \end{bmatrix}'$,

where $\mathbf{I_s}$ and $\mathbf{I_v}$ are identity matrices of station and village observations, respectively, $\mathbf{E_s}$ and $\mathbf{E_v}$ are the environmental covariance matrices (2x2) of the station traits (BW_s and EP_s) and the village traits (BW_v and EP_v), respectively.

For group observations, the same model as (5.2) was used, except that group averages of the 10 paternal-sibs were treated as if they were individual phenotypic records of each of the ten birds. This is an approximate approach described in Olson et al. (2006).

Selection for the reference breeding scheme was also based on breeding values estimated using model (5.2), except that the model was a bivariate model applied for 2 traits of BW_s and EP_s only. Selection for the reference scheme was indirect selection, in which selection index was:

$I = 0.078^* BW_s + 9.080^* EP_s$ (5.3)

However, (5.1) was still used as true breeding goal to assess genetic gain of all scenarios.

A combined measure of GxE, which represents the correlation between performances in the two environments, were based on the values of r_{g_BW} and r_{g_EP} with their economic indexes. Genetic correlation between (5.1) and (5.3) (r_{g_HI}) was calculated as:

$$r_{g\_HI} = \frac{Cov(H;I)}{\sqrt{Var(H) \times Var(I)}} \quad (5.4)$$

Where *Cov*(*H; I*) is genetic covariance between *H* and *I*; *Var*(*H*) is genetic variance of *H*; *Var*(*I*) is genetic variance of *I*.

Data analysis

For all scenarios, simulated output of total index genetic merit and the inbreeding coefficient from generation 5 to 20 were used for analyses. The index genetic merit of a scenario was the sum of true breeding values indexed with their economic values as in (5.1). The genetic merit of generation *t*, G_t, was the average of index true breeding values of all new-born individuals at generation *t*. Similarly, the inbreeding coefficient at generation *t*, F_t, was the average of inbreeding coefficients of individuals calculated by pedigree information.

For each replicate, genetic gain per generation (ΔG) was computed as the difference between G_{20} and G_5. The relative genetic gain per generation (*RG*) of VIO and VGO scenarios was calculated as the differences between their genetic gains and the mean of genetic gain of the corresponding reference scenarios divided by the mean of genetic gain of the reference scenarios.

$$RG = \frac{\Delta G_{alternative\ scenario} - Average\ \Delta G_{reference\ scenario}}{Average\ \Delta G_{reference\ scenario}} \times 100\%$$

Where *RG* is relative genetic gain per generation of VIO or VGO scenario over the reference scenario; $\Delta G_{alternative\ scenario}$ is genetic gain per generation of a replicate of VIO or VGO scenario; $Average\ \Delta G_{reference\ scenario}$

is the mean of genetic gain of 100 replicates of the reference scenario corresponding to the VIO or VGO scheme that had the same number of tested animals, and same r_{g_BW} and r_{g_EP}.

Rate of inbreeding per generation were computed as the negative of the slope of the regression of $ln(1-F_t)$ on t for F_5-F_{20} (Nirea et al., 2012).

Summary statistics for *RG* of VIO and VGO scenarios were based on 100 replicates. ANOVA were used to test direct and interaction effects of various factors on *RG.* The differences between scenarios were tested for significance using Tukey's HSD (honest significant difference, P <0.05). Summary statistics for rate of inbreeding of scenarios of VIO, VGO and reference schemes were also computed.

5.3. Results

The 4-way interaction of breeding scheme, r_{g_BW}, r_{g_EP} and number of tested animals were significant on *RG* with p <0.0001. As can be seen in Figure 5.2, all scenarios of VIO and VGO breeding schemes had genetic gain greater than the scenarios of the corresponding reference scheme. Relative genetic gains ranged from 21 to 268%.

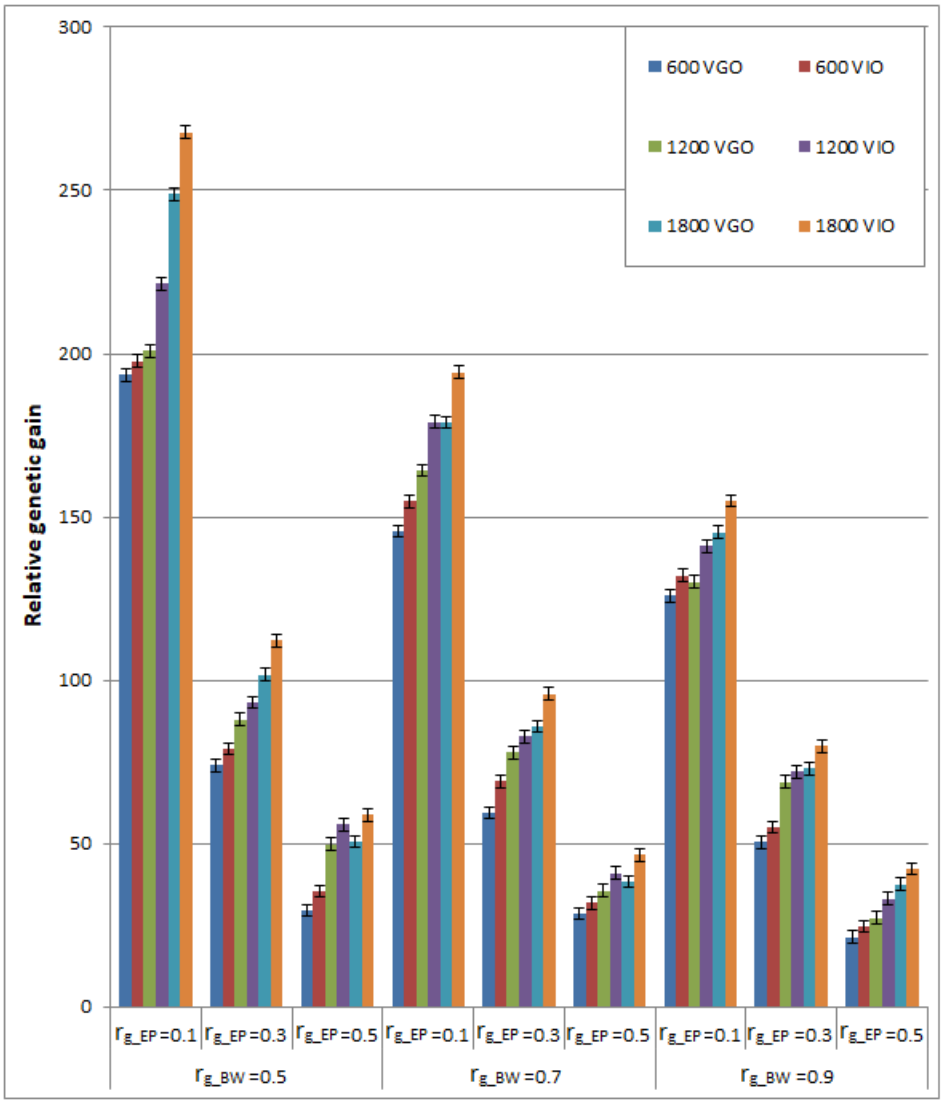


Figure 5.2: Means of relative genetic gains (%) (± SEM of 2%) of breeding scenarios with different genetic correlations between station and village bodyweight traits (r_{g_BW}) and egg production traits (r_{g_EP}) using either individual (VIO) or group recording (VGO) of 600, 1200 and 1800 village tested birds.

The addition of birds tested in the village condition increased *RG.* When number of village tested birds was 600, 1200 and 1800, on average, *RG* was 84, 98 and 112%, respectively.

The VIO breeding scheme had higher *RG* than VGO breeding scheme. On average, *RG* of VIO was 102% while it was 94% for VGO. With 600, 1200 and 1800 tested birds, *RG* of VGO were 81, 94 and 107%, respectively, and *RG* of VIO were 87, 102 and 117%, respectively. In all cases with the same number of village tested birds and the same correlations of r_{g_BW} and r_{g_EP}, *RG* of VGO was lower than that of VIO.

Lower genetic correlations between traits measured on station and village environments, lead to higher *RG.* With the maximum values for r_{g_EP} of 0.5 and r_{g_BW} of 0.9, on average, *RG* was 28% whereas *RG* was

225% for scenarios with the minimum values for r_{g_EP} of 0.1 and r_{g_BW} of 0.5. It seems that the magnitude of increase in *RG* was higher with a reduction in r_{g_EP} than with the reduction in r_{g_BW}. To have a better explanation of the trend of *RG,* r_{g_HI} can be used as an assessment of indirect selection of selection index I to the true breeding goal. *RG* decreased with increasing r_{g_HI} (Table 5.3).

Table 5.3: Mean of relative genetic gain (*RG* %) of breeding scenarios with different genetic correlations between station and village bodyweight traits (r_{g_BW}) and egg production traits (r_{g_EP}) corresponding to genetic correlations between breeding goal H and index I (r_{g_HI}). S.E.M. was 1%.

r_{g_BW}	r_{g_EP}	r_{g_HI}	Mean of RG
0.5	0.1	0.16	222
0.7	0.1	0.20	170
0.9	0.1	0.23	138
0.5	0.3	0.33	91
0.7	0.3	0.36	78
0.9	0.3	0.40	67
0.5	0.5	0.50	47
0.7	0.5	0.53	37
0.9	0.5	0.56	31

Genetic gains per generation of VIO and VGO breeding scenarios can be seen in Table 5.4. Similar to *RG, ΔG* of VIO and VGO scenarios increased with increasing number of village tested animals. Genetic gains of VIO scenarios were also higher than that of VGO scenarios. However, unlike *RG,* lower genetic correlations between traits measured in station and village environments (r_{g_EP} and r_{g_BW}) resulted in lower *ΔG*. Meanwhile, *ΔG* increases with a higher genetic correlation between environments (r_{g_HI}), but at a lower relative increase as can be seen in Table 5.4.

Table 5.4: Mean of genetic gains per generation (*ΔG*) (± SD) of breeding scenarios with different genetic correlations between station and village bodyweight traits (r_{g_BW}) and egg production traits (r_{g_EP}) using either individual (VIO) or group recording (VGO) of village observations of 600, 1200 and 1800 birds.

r_{g_BW}	r_{g_EP}	r_{g_HI}	600		1200		1800	
			VIO	VGO	VIO	VGO	VIO	VGO
0.5	0.1	0.16	20.4 ± 2.4	20.1± 2.7	23.7 ± 2.1	22.2 ± 2.5	25.9 ± 1.8	24.5 ± 2.0
0.7	0.1	0.20	21.9 ± 2.1	21.2 ± 2.2	25.1 ± 2.3	23.7 ± 2.1	26.6 ± 2.0	25.2 ± 2.2
0.9	0.1	0.23	23.7 ± 2.3	23.0 ± 2.1	26.5 ± 2.1	25.3 ± 2.0	28.0 ± 2.0	26.9 ± 2.3
0.5	0.3	0.33	21.1 ± 2.2	20.5 ± 2.2	24.0 ± 2.1	23.4 ± 2.3	25.9 ± 2.2	24.7 ± 2.3
0.7	0.3	0.36	22.8 ± 2.3	21.5 ± 2.3	25.3 ± 2.2	24.6 ± 2.2	27.3 ± 2.1	25.9 ± 2.0
0.9	0.3	0.40	24.7 ± 2.3	24.0 ± 2.3	26.6 ± 2.0	26.2 ± 1.6	28.2 ± 1.8	27.1 ± 1.9
0.5	0.5	0.50	23.0± 2.1	22.0 ± 2.4	25.7 ± 2.2	24.7 ± 2.5	27.4 ± 2.1	26.0 ± 2.1
0.7	0.5	0.53	24.0 ± 2.4	23.4 ± 2.3	26.7 ± 2.2	25.7 ± 2.1	28.1 ± 2.5	26.6 ± 2.1
0.9	0.5	0.56	25.8 ± 2.1	25.1 ± 2.2	27.7 ± 2.0	26.5 ± 1.7	28.9 ± 1.9	28.0 ± 2.1

The rates of inbreeding decreased as number of tested animals increased. They, on average, were 2.00, 1.99 and 1.97% for scenarios with 600, 900 and 1800 tested animals, respectively. Higher rates of inbreeding were found in the VGO scenarios than in the VIO scenarios. The rates of inbreeding, on average, were 1.70% for the reference scenarios, 2.10% for VIO scenarios and 2.15% for VGO scenarios. The rates of inbreeding had a reducing tendency as r_{g_EP} and r_{g_BW} increased.

5.4. Discussion

In this study, breeding schemes for village dual-purpose poultry using group and individual recordings of village and station performances at different levels of GxE interaction were compared. Results showed that village observations significantly improved genetic gains of VIO and VGO compared to the reference breeding scheme. The improvement was larger in VIO than in VGO. Increasing number of village tested birds also increased genetic gain. Higher genetic correlations between traits measured in station and village environments lead to lower relative genetic gain, but higher absolute genetic gain.

Effects of village observation on genetic gain

Increasing the number of animals tested in village improved accuracy of selection, and thus genetic gains of VIO and VGO schemes compared to the reference schemes. The main difference between the reference breeding scheme and its alternatives was the type of tested information. Village observations were direct

phenotypes while station observations were correlated phenotypes. The reference scheme had only station performance while its alternatives had both station and village performances. Selection in the reference scheme is an indirect selection approach, and therefore, it results in the lowest accuracy of selection compared to its alternatives.

The value of village observations increased when the genetic correlation between station and village traits was lower. The scenarios with r_{g_BW} of 0.9 and r_{g_EP} of 0.5 gave lower relative genetic gain than those with r_{g_BW} of 0.5 and r_{g_EP} of 0.1. Nonetheless, the absolute genetic gain was larger in scenarios with higher genetic correlations as contribution of both station and village observations to accuracy of selection increased.

It has been suggested that if the genetic correlation between performance in the selection and production environments is less than 0.8, breeding program with information from the production environment would be worthwhile to improve genetic gain (Robertson, 1959). Other studies have also shown that a significantly higher genetic gain can be achieved with performance information from the production environment (Bijma & Arendonk, 1998; Mulder & Bijma, 2005). However, when genetic correlation between the performance in selection and production environments is high, for example 0.9, a large number of animals need to be tested in the production environment for a significant improvement in genetic gain.

In our study, relative genetic gains were positive in all scenarios of VIO and VGO with any number of birds tested in village, r_{g_BW} or r_{g_EP}. In other studies, to model GxE, a single trait in two environments is often used (Bijma & Arendonk, 1998; Mulder & Bijma, 2005). To be comparable to other studies, instead of r_{g_BW} and r_{g_EP}, r_{g_HI} should be used as a representative of genetic correlation between station and village environments. It takes into account the variances and covariances of BW and EP traits measured in the two environments with their economic indexes. The value of r_{g_HI} reflects the magnitude of indirect selection on the selection index I to the true breeding goal. It describes the extent of GxE when more than one trait is measured in two environments. In the simulation, r_{g_HI} was 0.16-0.56, which might explain the high relative genetic gains of all VIO and VGO scenarios.

It was expected that both the increases of relative genetic gain and decreases in absolute genetic gain would correspond to increases of r_{g_HI}. However, an increasing tendency of absolute genetic gain did not correspond to the increase of r_{g_HI} (Table 5.4). Possible explanations may include the two-stage selection for BW in males and that EP is a sex limited trait (50% fewer records for EP than for BW), thus a change of r_{g_EP} has a different impact on absolute genetic gain than a change of r_{g_BW}.

Group versus individual observation

VGO breeding scheme was similar to VIO, except that recordings of village performance were in groups of 10 paternal-sibs. Our findings showed that VGO had lower *RG* than VIO, which is due to a lower accuracy of prediction of breeding values using group recording. Pooling birds in groups reduced the amount of information that was provided for each individual. Nonetheless, VGO had substantially increased genetic gains compared to the reference breeding scheme and reduction of the absolute genetic gain in comparison to the corresponding VIO scenario was at most 6% (Table 5.4.).

Other studies have analysed pooled data, in which pooled observations were groups of random animals, full-sibs, half-sibs and descendants of maternal grand sire (Biscarini et al., 2008; Nurgiartiningsih et al., 2004; Olson et al., 2006; Peeters et al., 2013). From these studies, it can be concluded that estimation of breeding values from pooled data is theoretically and practically feasible for selection, particularly when the pooled observations are groups of closely related animals.

Biscarini et al. (2008) illustrated that correlations between EBV based on individual observation and the pooled observation of 4 half-sib animals were 0.703-0.748 for EBV of the own animals, 0.814-0.891 for EBV of their sires with more than 10 offspring and 0.847-0.880 for EBV of their dams with more than 4 offspring. Nurgiartiningsih et al. (2004) also demonstrated high correlations between EBV based on individual and group observations which were, on average, 0.844 for EBV of the animals and 0.943 for EBV of their sires. Olson et al. (2006) studied accuracies of predicting breeding values from individual and group observation using simulation. They found that in the absence of pen effects, accuracies of EBV of animals themselves or their sires would be improved when animals allocated in a group were more related and when size of each group was smaller given the same total number of animals.

In our study, to estimate EBV, selection candidates of VGO scenarios could have indirect information from individual observations of the correlated traits (BW_s and EP_s) of their own and parents' performance and direct information from pooled observations of the desired village traits (BW_v and EP_v) of their sibs. The pooled observations were groups of birds that had half-sib and full-sib relationship to the selection candidates. By averaging observations of the sib mixture, effects of dams mated to a sire on their offspring cannot be distinguished. The pooled observations can be only approximated as average of half-sibs. Meanwhile, effects of dams, full-sib and half-sib relationships can be taken into account in predicting EBV of selection candidates in VIO scenarios, which resulted in a higher genetic gain in VIO than in VGO scenarios.

Nonetheless, the differences between accuracy of selection of VIO and VGO were not substantial. With r_{g_BW} of 0.5 and r_{g_EP} of 0.1, accuracy of EBVs of selection candidates was 0.863, 0.892 and 0.917 for VGO

with 600, 1200 and 1800 village tested birds, respectively, while the accuracy of EBVs was 0.868, 0.908 and 0.925 for VIO with 600, 1200 and 1800 village tested birds, respectively. With r_{g_BW} of 0.9 and r_{g_EP} of 0.5, the accuracy of EBVs was 0.917, 0.927 and 0.935 for VGO with 600, 1200 and 1800 village tested birds, respectively; and 0.920, 0.934 and 0.940 for VIO with 600, 1200 and 1800 village tested birds, respectively.

Methodology

In our study, high relative genetic gains were achieved for VIO and VGO scenarios, and none of their replicates had negative relative genetic gains. This is due to 3 important assumptions including strong GxE, unchanged number of selection candidates and no common maternal effects.

GxE was modelled for BW at r_{g_BW} of 0.5-0.9 and EP at r_{g_EP} of 0.1-0.5, which represents quite strong interactions. Conventional breeding programs are usually carried out under conditions most favourable for the expression of genotypes. One of the important reasons for this is that GxE is often small, especially for commercial breeds where production animals are reared in enclosed, highly controlled conditions, similar to the station situation. However, the differences between village and breeding station are likely to be more substantial. Therefore, if birds are selected under station conditions of sufficient and balanced diets, absence of infectious diseases and minimum of stress, strong GxE will be expected.

Number of selection candidates was assumed to be unchanged, even for the reference breeding scheme in which tested birds were assumed to have station observations. This assumption is not reasonable in practice, but it was included to quantify benefit of village observations. In theory, as long as genetic correlation between traits measured in station and village environments is less than 1, village observations would provide additional genetic gains for VIO and VGO. Alternatively, if the combined number of birds for selection and village testing was constant, the use of birds for village testing in VIO schemes would not be beneficial for genetic gain with r_{g_HI} above 0.8 due to reduced selection intensity (Chu et al., 2018; Mulder & Bijma, 2005; Robertson, 1959). The use of birds for village testing in VGO schemes would only become beneficial when r_{g_HI} was even lower than the r_{g_HI} of VIO schemes.

Common maternal effects were not included in our simulation. The inclusion of the common maternal effects would have relatively slight effects on genetic gain of VIO if birds from different families are randomly distributed to smallholders. In contrast, it would reduce considerable genetic gains of VGO as members of the group with pooled observations were paternal-sibs. However, the common maternal effects are negligible for the traits of selection in breeding program for village poultry. The traits for selection are often at relatively old age, for example, BW at 16 or 20 weeks of age and EP at 40 or 44 weeks of age. At these ages, common maternal effects for BW and EP would be insignificant. Common maternal effects for BW reduce as birds age (Begli et al., 2016; Dana et al., 2011; Prado-Gonzalez et al.,

2003). The dam effects of BW disappeared at 8 weeks of age (Prado-Gonzalez et al., 2003) and 12 weeks of age (Dana et al., 2011). Common dam effects are usually not included in the model for EP traits as they are expressed late in bird life.

Application of breeding schemes for village poultry production

Poultry breeding for village production by poor and nutritionally insecure people in the rural and peri-urban regions of the Sub-Saharan Africa must accept the reality that people prefer dual-purpose chicken in a scavenging or semi-scavenging system (Dana et al., 2010). High investment for commercial housing shed, supplementation of feed and expanded flock size can lead to unsteady net returns. Such a risky investment was one of the main reasons that village farmers were reluctant to spend on the inputs (Wondmeneh, 2015). It is shown that the use of the scavenging production system for smallholders brings in higher economic returns than the use of the semi-intensive or intensive system (FAO, 2010). Therefore, to improve the livelihood of the targeted people, a proper breeding program for village poultry production is required.

Using village observations, breeding schemes VIO and VGO would be appropriate for improving genetic gain of a breeding program and possibly maintaining adaptability traits which are major advantages of indigenous chicken in village production. However, implementation of VIO requires individual records of pedigree and measurement of phenotypes under village conditions. Routine recording phenotypes for individual birds is most likely not possible in village production systems. Measurement of individual phenotypes by smallholder farmers often has low accuracy (Lwelamira, 2012). Implementation of VGO is simpler in practice compared to VIO. Although lower genetic gain is predicted for VGO, the increased accuracy of data recording in VGO may make up for this. Group recording in the VGO breeding scheme reduces the complexity of tracing and recording process. Therefore, the recommended breeding scheme for village poultry production is VGO. Testing 600 birds in the village environment results in significant genetic gain for the program, compared to testing them on station.

5.5. Conclusions

Village observations significantly increased genetic gain compared to station observations. The improvement was only slightly larger with individual observations (VIO) than with group observations (VGO). Higher genetic correlations between traits measured in station and village environments led to higher genetic gain, but lower relative genetic gain in VIO and VGO scenarios. In assessing relative genetic gains from village observations for a breeding program in presence of GxE, r_{g_HI}, the genetic correlation between station and village breeding objective, should be used to model GxE as it explained better the

magnitude of GxE than r_{g_BW} or r_{g_EP} alone. Breeding schemes that use village group recording are applicable for breeding indigenous dual-purpose poultry where a strong GxE is expected.

References:

Begli, H. E., Torshizi, R. V., Masoudi, A. A., Ehsani, A., & Jensen, J. (2016). Longitudinal analysis of body weight, feed intake and residual feed intake in F_2 chickens. *Livestock Science,* 184, 28-34.

Bekele, F., Gjoen, H. M., Kathle, J., Adnoy, T., & Abebe, G. (2009). Genotype x environment interaction in two breeds of chickens kept under two management systems in Southern Ethiopia. *Tropical Animal Health and Production,* 41, 1101-1114.

Bijma, P., & Arendonk, J. A. M. (1998). Maximizing genetic gain for the sire line of a crossbreeding scheme utilizing both purebred and crossbred information. *Animal Science,* 66, 529-542.

Biscarini, F., Bovenhuis, H., & van Arendonk, J. A. (2008). Estimation of variance components and prediction of breeding values using pooled data. *Journal of animal science,* 86, 2845-2852.

Cahaner, A., Deeb, N., & Gutman, M. (1993). Effect of the plumage-reducing naked-neck (Na) gene on the performance of fast-growing broilers at normal and high ambient temperatures. *Poultry science,* 72.

Chen, C.F., Gourichon, D., Huang, N.Z., Lee, Y.P., Bordas, A., & Tixier-Boichard, M. (2009). Performance comparison of dwarf laying hens segregating for the naked neck gene in temperate and subtropical environments. *Genetics Selection Evolution,* 41, 1-11.

Chu, T.T., Alemu, S.W., Norberg, E., Sørensen, A.C., Henshall, J., Hawken, R., & Jensen, J. (2018) Benefits of testing birds in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken. Manuscript submitted for publication.

Dana, N., van der Waaij, L. H., Dessie, T., & van Arendonk, J. A. M. (2010). Production objectives and trait preferences of village poultry producers of Ethiopia: implications for designing breeding schemes utilizing indigenous chicken genetic resources. *Tropical Animal Health and Production,* 42, 1519-1529.

Dana, N., vander Waaij, E. H., & van Arendonk, J. A. M. (2011). Genetic and phenotypic parameter estimates for body weights and egg production in Horro chicken of Ethiopia. *Tropical Animal Health and Production,* 43, 21-28.

Dessie, T., Alemu, Y., & Peters, K. (2000). Indigenous chickens in Ethiopia: genetic potential and attempts at improvement. *World's Poultry Science Journal,* 56, 45-54.

FAO. (2010). Chicken genetic resources used in smallholder production systems and opportunities for their development. In P. Sørensen (Ed.), *FAO Smallholder Poultry Production Paper No. 5*. Rome.

Horst, P. (1985). *Effects of genotype × environment interactions on efficiency of improvement of egg production.* Paper presented at the Proceedings of the 18th Poultry Science Symposium and of the 25th British Poultry Breeders Round Table.

Kapell, D. N., Hill, W. G., Neeteson, A. M., McAdam, J., Koerhuis, A. N., & Avendano, S. (2012). Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments. *Poultry science,* 91, 565-574.

Lwelamira, J. (2012). Genotype-Environment (GXE) Interaction for Body Weights for Kuchi Chicken Ecotype of Tanzania Reared Under Intensive and Extensive Management. *Global Journal of Medical Research,* 12*, 51-57*.

Lwelamira, J., Kifaro, G. C., & Gwakisa, P. S. (2009). Genetic parameters for body weights, egg traits and antibody response against Newcastle Disease Virus (NDV) vaccine among two Tanzania chicken ecotypes. *Tropical Animal Health and Production,* 41, 51-59.

Mathur, P. K. (2003). Genotype-environment interactions: problems associated with selection for increased production. In W. M. Muir & S. E. Aggrey (Eds.), *Poultry Genetics, Breeding and Biotechnology* (pp. 83-99): CAB International.

Mathur, P. K., & Horst, P. (1994). Methods for evaluating genotype × environment interactions illustrated with laying hens. *Journal of Animal Breeding and Genetics,* 111.

Mulder, H., & Bijma, P. (2005). Effects of genotype× environment interaction on genetic gain in breeding programs. *Journal of animal science,* 83, 49-61.

N'Dri, A. L., Sellier, N., Tixier-Boichard, M., Beaumont, C., & Mignon-Grasteau, S. (2007). Genotype by environment interactions in relation to growth traits in slow growing chickens. *Genetics Selection Evolution,* 39, 513-528.

Niknafs, S., Nejati-Javaremi, A., Mehrabani-Yeganeh, H., & Fatemi, S. A. (2012). Estimation of genetic parameters for body weight and egg production traits in Mazandaran native chicken. *Tropical Animal Health and Production,* 44, 1437-1443.

Nirea, K. G., Sonesson, A. K., Woolliams, J. A., & Meuwissen, T. H. (2012). Effect of non-random mating on genomic and BLUP selection schemes. *Genetics Selection Evolution,* 44, 11.

Nurgiartiningsih, V. M., Mielenz, N., Preisinger, R., Schmutz, M., & Schueler, L. (2004). Estimation of genetic parameters based on individual and group mean records in laying hens. *British Poultry Science,* 45, 604-610.

Okeno, T. O., Magothe, T. M., Kahi, A. K., & Peters, K. J. (2012). Application of risk-rated profit model functions in estimation of economic values for indigenous chicken breeding. *Tropical Animal Health and Production,* 44, 1279-1287.

Okeno, T. O., Magothe, T. M., Kahi, A. K., & Peters, K. J. (2013). Breeding objectives for indigenous chicken: model development and application to different production systems. *Tropical Animal Health and Production,* 45, 193-203.

Oleforuh-Okoleh, V. U. (2011). Estimation of genetic parameters and selection for egg production traits in a Nigerian local chicken ecotype. *ARPN Journal of Agricultural and Biological Science,* 6, 12.

Olson, K. M., Garrick, D. J., & Enns, R. M. (2006). Predicting breeding values and accuracies from group in comparison to individual observations. *Journal of animal science,* 84, 88-92.

Pedersen, L., Sørensen, A., Henryon, M., Ansari-Mahyari, S., & Berg, P. (2009). ADAM: A computer program to simulate selective breeding schemes for animals. *Livestock Science,* 121, 343-344.

Peeters, K., Ellen, E. D., & Bijma, P. (2013). Using pooled data to estimate variance components and breeding values for traits affected by social interactions. *Genetics Selection Evolution,* 45*, 27*.

Prado-Gonzalez, E., Ramirez-Avila, L., & Segura-Correa, J. (2003). Genetic parameters for body weights of Creole chickens from Southeastern Mexico using an animal model. *Livestock Research for Rural Development,* 15, 27-31.

Robertson, A. (1959). The sampling variance of the genetic correlation coefficient. *Biometrics,* 15, 469-485.

Wolc, A., Arango, J., Settar, P., Fulton, J., O'Sullivan, N., Preisinger, R., . . . Dekkers, J. (2013). Analysis of egg production in layer chickens using a random regression model with genomic relationships. *Poultry science,* 92, 1486-1491.

Wondmeneh, E. (2015). *Genetic improvement in indigenous chicken of Ethiopia.* (PhD Thesis), Wageningen University, Netherlands.

Wondmeneh, E., van Der Waaij, E., Dessie, T., Udo, H., & Van Arendonk, J. (2015). Adoption of exotic chicken breeds by rural poultry keepers in Ethiopia. *Acta Agriculturae Scandinavica, Section A—Animal Science,* 64, 210-216.

6

Optimized grouping to increase accuracy of prediction of breeding values based on group records in genomic selection breeding programs

Thinh T. Chu[1,2*], John W. M. Bastiaansen[2], Peer Berg[1,3], Hans Komen[2]

[1] Aarhus University, Center for Quantitative Genetics and Genomics, 8830 Tjele, Denmark

[2] Wageningen University & Research, Animal Breeding and Genomics, 6709 PG Wageningen, the Netherlands

[3] Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, 1432 Aas, Norway

Corresponding author's email: chu.thinh@mbg.au.dk

GSE accepted

DOI: 10.1186/s12711-019-0509-z

Abstract

Background: Accuracy of prediction of breeding values from phenotypic group records depends on group structure. When genotyping information is available before phenotyping, utilization of this genotyping information to form groups may improve accuracy of prediction from group records. This study analyzed two grouping methods based on genomic information: unsupervised clustering implemented in STRUCTURE software and supervised clustering based on genomic relationships.

Results: Genetic variances estimated with GBLUP models from group records were consistent with estimates from individual records. Using genomic information to make groups led to higher genomic relationship between group members than random grouping of paternal half-sibs and random grouping of full-sibs. Genomic relationships between group members that were formed from the supervised clustering method depended on group sizes, number of groups, family sizes, genome sizes and number of surplus genotyped offspring. The grouping methods based on genomic information resulted in higher accuracy of GEBV prediction (1.2-1.7%) compared to random grouping of full-sibs and (11.1-11-8%) random grouping of paternal half-sibs.

In addition, grouping methods based on genomic information led to lower coancestry coefficients between top ranking GEBV animals. Of the two proposed methods, supervised clustering based on genomic relationships was superior in term of computation requirements, applicability and accuracy of GEBV prediction. Benefits of the supervised clustering method could be were further elaborated for accuracy of GEBV estimated from group records when there were surplus offspring that were available for the grouping procedures. In this situation, the advantage of the supervised clustering method was up to 4.5% compared to random grouping of full-sibs and 14.7% compared to random grouping of paternal half-sibs.

Conclusions: The use of genotyping information for grouping gives additional accuracy of selection when phenotypic group records are used in genomic selection breeding programs.

Keywords: genomic; group records; clustering.

6.1. Introduction

Obtaining continuous individual records is difficult and expensive for economically important traits such as feed efficiency and egg production. For breeding programs to improve genetic gain in the presence of GxE, continuous individual recording for these traits can even be impossible for animals tested in commercial production environments or in village conditions. In such situations, pooled data on group records can be an alternative. The use of pooled data was shown to be feasible for predicting variance components and breeding values of animals with pedigree-based BLUP (Biscarini *et al.*, 2008; Olson *et al.*, 2006; Peeters *et al.*, 2013; Su *et al.*, 2018). These studies showed that accuracy of prediction depends on relationship between group members. Predictions became more accurate when group members were more closely related. In the additive numerator relationship matrix, relationship coefficients between members of a full-sib group are all the same, 0.5 for unrelated parents or higher for inbred parents. Genomic information gives a better measure of the relationships between animals than pedigree information. The realized genomic relationships between specific pairs of full-sibs is known to vary with the standard deviation depending on genome size and number of chromosomes (VanRaden, 2007). Empirical values show a range from 0.27-0.70 for chicken (Lourenco *et al.*, 2015b) and 0.35-0.65 for cattle (Calus *et al.*, 2011). We hypothesized that genomic information could be exploited to improve accuracy of prediction from group records when genotyping data would be available before animals are grouped for phenotype testing.

Accuracy of prediction from group records increases with increasing relationships between animals within groups (Olson *et al.*, 2006; Peeters *et al.*, 2013; Su *et al.*, 2018), thus increase in genomic similarity of group members may improve the accuracy. Unsupervised clustering of genetically similar individuals into groups based on genomic data is implemented in a program named STRUCTURE (Pritchard *et al.*, 2000). This Bayesian, model-based program is used widely in analysis of population structure (Pritchard *et al.*, 2000). The program integrates over the parameter space, infers population structure and makes cluster assignments for every individual (Pritchard *et al.*, 2000). The number of subpopulations, or clusters, can be given or estimated. An output of the program is the membership coefficient or probability that an individual belongs to a given cluster. Clustering individuals into subpopulations is a similar concept to clustering animals that have close relationships into the same groups. However, this approach may not be optimal for designing breeding programs because the number of individuals assigned to each cluster can vary under unsupervised clustering. For example, the breeding facilities, which are usually fixed system, can accommodate 4 groups with 4 animals per group, but unsupervised clustering ends up with animals

clustered into only 3 groups. Besides, the membership coefficient of an animal from STRUCTURE does not always show a clear distinction of which group the animal belongs to.

In addition to the grouping method based on the STRUCTURE program, we therefore propose a grouping method that maximizes the relationships between animals within groups, based on the realized genomic relationship matrix. This grouping method is a supervised clustering approach, in which number of groups and group sizes are defined as fixed input variables. The method does not use genotyping data directly, but indirectly through the realized genomic relationship matrix.

The objectives of this study were to (1) compare both grouping methods based on genotyping information to improve accuracy of selection from group records, (2) use GBLUP models to estimate variance components from group records, and (3) investigate effects of surplus genotyped offspring, number of groups, family sizes and genome sizes on breeding schemes that used the proposed grouping method.

6.2. Methods

Simulation of a population of animals was implemented with the following steps: (1) simulation of individual genotype and trait records using the stochastic simulation program ADAM (Pedersen *et al.*, 2009), (2) allocation of simulated animals into groups based on pedigree or genomic information using different grouping methods, and (3) simulating phenotypic group records. Therefore, animals were simulated to be genotyped, grouped and then phenotyped. After that, variance components and breeding values were estimated from individual records or from group sum records using genomic BLUP model.

Breeding schemes, genotype and trait simulations

The historical base populations with genomic structure were from Chu *et al.* (2018) in which the simulated genome consisted of 26 chromosomes with a total length of 916 cM. Segregating loci of 2k QTL and 40k neutral markers that were randomly distributed along the genome were used for trait simulations and genotyping information, respectively. The segregating loci of QTL and markers had a minor allele frequency of at least 0.05 in the base population. Inheritance of QTL and markers from parents to descendants followed the standard principles of Mendelian inheritance, and allowed for recombination as described in Chu *et al.* (2018). From the base population, 20 sires and 200 dams were used for hierarchical mating scheme where one sire is mated with ten dams, but each dam mated with a single sire only. Each dam produced 16 offspring, thus the total number of offspring was 3200. Sex was randomly assigned to the offspring with ratio 1:1. Only one generation of offspring was simulated. All sires, dams and offspring had genotyping information.

True breeding value of each individual was the sum of its QTL effects. Allele substitution effects of QTL were randomly sampled from a normal distribution N[0, 1], and then rescaled to achieve the initial additive genetic variance of 0.3 in the base population. Simulated phenotype of individual records was the sum of the true breeding value and an environmental deviation term: $y_i = \mu + tbv_i + e_i$,where y_i is the individual phenotypic record of animal *i*; μ is mean of the trait equal to 0; true breeding value and residual environmental deviation of animal *i* are tbv_i and e_i, respectively. The residual environmental deviations were drawn from normal distribution N[0, 0.7]. Thus, the phenotypic variance was 1.

Four animals were pooled in a group, making 800 groups. For individual records, each animal had its own records. For group records, only the sum of simulated individual phenotypes from the four animals in a group was available.

Estimation of variance components and prediction of breeding values

Individual records and group records were used to estimate variance components and GEBV using the DMUAI module from the DMU software package (Madsen and Jensen, 2013). The model for individual records in matrix notations was:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e} \qquad (\text{GBLUP}_i)$$

where **y** is a vector of phenotypic individual records; μ is the mean; **g** and **e** are vectors of genomic breeding values (GEBV) and residuals, respectively. These vectors are assumed to be normally distributed: $\mathbf{g} \sim N[0, \mathbf{G}\,\sigma^2_g]$ and $\mathbf{e} \sim N[0, \mathbf{I}\,\sigma^2_e]$, where **G** is genomic relationship matrix constructed from marker data; **I** is an identity matrix associating residuals to individual phenotypic records; σ^2_g and σ^2_e, are additive genetic variance and residual variance, respectively. The incidence matrix **Z** associates **g** to individual phenotypic records.

When group records were analyzed, the model for estimation of variance components and prediction of GEBV was similar to the exact model in Olson *et al.* (2006) and the model in Su *et al.* (2018) except that realized genomic relationship matrix was used instead of additive numerator relationship matrix. The models applied to equal group sizes was:

$$\mathbf{y}^* = \mathbf{1}\mu + \mathbf{Z}^*\mathbf{g} + \mathbf{e}^* \qquad (\text{GBLUP}_{gr})$$

where $\mathbf{y}^*$ is a vector of group records with number of elements equal to number of groups; **g** is a vector of GEBV as described above in the models for individual records: $\mathbf{g} \sim N[0, \mathbf{G}\,\sigma^2_g]$; $\mathbf{e}^*$ is a vector of residuals: $\mathbf{e}^* \sim N[0, \mathbf{R}\,\sigma^2_e]$, where **R** is a diagonal matrix and diagonal elements are equal to group size. Matrix $\mathbf{Z}^*$ is an incidence matrix associating **g** to phenotypic group records. Matrices of **Z** and $\mathbf{Z}^*$ have equal number of

columns, but **Z** and **Z*** have a number of rows equal to number of individual records and group records, respectively.

For example, eight animals 3-10 that were offspring of animals 1 and 2 were grouped in two groups of four animals. Phenotypic records of groups were 2.6 and 3.5. The model for group records is:

$$\begin{bmatrix}2.6\\3.5\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}\mu + \begin{bmatrix}0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}g_1\\g_2\\g_3\\\dots\\g_{10}\end{bmatrix} + \begin{bmatrix}\sum_{i=3}^{6} e_i\\\sum_{i=7}^{10} e_i\end{bmatrix}$$

The realized genomic relationship matrix **G** was constructed from marker data of all sires, dams and offspring individuals using VanRaden (2008) method 1:

$$\mathbf{G} = \frac{\mathbf{MM}'}{2\sum p_j(1-p_j)}$$

where **M** is a matrix which has number of rows equal to number of animals and number of columns equal to number of markers; matrix M is centered so that elements in column *j* are 0-2p_j, 1-2p_j and 2-2p_j for genotypes A_1A_1, A_1A_2 and A_2A_2, respectively; p_j is allele frequency of A_2 at locus *j* computed from the marker data of all dams, sires and offspring. Division by $2\sum p_j(1-p_j)$ scales matrix **G** to be analogous to the pedigree-based numerator relationship matrix. The realized genomic relationship matrix **G** was used for genetic evaluation models, grouping methods and investigation of relationship distributions.

Grouping methods

For group records, animals were pooled into a group based on either pedigree or genomic information. Grouping methods based on pedigree information were random grouping of full-sibs and random grouping of paternal half-sibs. Grouping methods based on genomic information were unsupervised clustering based on genotypes and supervised clustering based on the genomic relationship matrix.

Random grouping of paternal half-sibs: Animal allocation into groups was based on having a common sire. Four animals that were paternal half-sibs were randomly pooled into one group.

Random grouping of full-sibs: Animal allocation into groups was based on having a common sire and dam. Four animals from the same full-sib family were randomly pooled into one group.

Unsupervised clustering based on genotypes: When unsupervised clustering analysis using STRUCTURE program was applied to all 3200 animals, paternal half-sibs from a sire were always clustered into one group even if assumed number of clusters was set at 800. When the unsupervised clustering analysis was applied to a group of paternal half-sibs from a single sire, full-sibs from a family were always clustered into one group. Therefore, clustering analysis was carried out separately for every full-sib family of 16

animals. Admixture model in STRUCTURE was used (Pritchard *et al.*, 2000). Number of clusters was set to be 4, and allele frequencies were assumed to be correlated among clusters (Falush *et al.*, 2003). Cluster membership coefficients of the 16 animals from the output was used for group allocations. Animals were pre-allocated to 4 groups based on the highest membership coefficients of these animals. In many cases, the pre-allocated 4 groups did not all have the expected number of 4 animals. Four animals with the top ranking of membership coefficients from the biggest group were allocated to the first group. The remaining 12 animals were then pre-allocated to 3 groups based on the highest membership coefficients of these animals. Four animals with the top ranking of membership coefficients from the biggest group were allocated to the second group. Similarly, 4 animals were allocated to the third group and the fourth group consisted of the remaining 4 animals. These grouping allocation procedures were applied to all 200 full-sib families to make 800 groups in total. By the unsupervised clustering method based on genotypes, animals within a group were always full-sibs.

Supervised clustering based on genomic relationships: A supervised clustering method scripted in R (R Core Team, 2018) was developed to pool 4 animals into groups based on realized genomic relationships between animals. Applying this approach to all 3200 offspring was time-consuming because of the many possibilities for allocating the offspring into 800 groups. The probability that half-sibs or non-related animals would be placed in the same group was extremely low. Therefore, grouping was carried out separately for every full-sib family of 16 animals in an evolutionary algorithm as follows:

- Animals from a full-sib family were randomly assigned to 4 groups with 4 animals in each group.
- An exchange of two randomly chosen animals between two randomly chosen groups was proposed.
- Group membership was updated if the proposal resulted in an increase of the mean genomic relationship between members within groups.
- The iteration was considered converged when the exchange of two animals between two groups did not increase the mean of genomic relationships for a certain number of iterations.

The exchange of two animals between two groups was a random process, but if the exchange did not increase genomic relationships between group members, these two animals would not be chosen for the exchange in the next iteration until a new set of groups of animals was formed. Therefore, the number of iterations without the changes in genomic relationships were $n_a^2 \times \sum_{i=1}^{n_g-1} i$, where na is number of animals per group (group size) and ng is number of groups per full-sib family. The numbers of iterations are the possibilities of forming a new set of groups of animals when two animals are randomly chosen from two random groups. In addition, the exchange of two animals is a conditional event given that a

certain set of groups of animals has formed. Therefore, the evolutionary algorithm above was re-run for 300 times that formed up to 300 different sets of groups. The set of groups of animals that gave the highest genomic relationships between group members was chosen. The number 300 is an empirical number after different trials to get a set of groups of animals with the highest genomic relationships between group members.

As a result of this supervised clustering method based on genomic relationships the offspring were allocated into 800 different groups within each of which animals were always full-sibs.

Sensitivity analysis

To investigate effects of surplus genotyped offspring, number of groups, family sizes and genome sizes on breeding schemes that used random grouping of paternal half-sib, random grouping of full-sibs and supervised clustering method based on genomic relationships, four extra simulations were carried out. Group sizes and number of group per full-sib family in sensitivity analysis simulation 1 (SS1) were the same as in the main study, but the family sizes were varied at 32 and 48 offspring per full-sib family (Table 6.1). In SS1, all offspring, which had genotypes, were used in grouping procedures, but after grouping, some surplus offspring were not assigned to any groups or tested for phenotyping. For comparisons, individual records of SS1 were obtained from only 16 offspring per dam that were randomly chosen from each full-sib family. Sensitivity analysis simulation 2 (SS2) had the same breeding structure as in the main study, but group sizes and number of groups per full-sib family were varied. In sensitivity analysis simulation 3 (SS3), number of groups was constant, but family size (number of offspring per dam) was varied. Sensitivity analysis simulation 4 (SS4) was the same as in the main study, except that the simulated genome consisted of 30 chromosomes with the length of 100cM for each. The total length of the genome was 3000cM.

Table 6.1. Group sizes, number of groups and family sizes for sensitivity simulation (SS) 1-4.

Investigated factors	Main study	SS1	SS2	SS3	SS4
Number of groups per full-sib family	4	4	8; 2	4	4
Family sizes (offspring per dam)	16	32; 48	16	8; 32; 48	16
Surplus genotyped offspring without phenotypes per full-sib family	0	16; 32	0	0	0
Group sizes (animals per group)	4	4	2; 8	2; 8; 12	4
Genome size	916 cM	916 cM	916 cM	916 cM	3000 cM

Supervised clustering method based on genomic relationships in SS1-4 was similar as in the main study with the aim to maximize relationships between animals within groups. With SS1, one extra group was

added that included all surplus animals. The probability of sampling groups for the exchange between two random animals was corresponding to number of animals of these groups.

Data analysis

Scenarios were replicated 100 times. Accuracy of GEBV predictions was computed as correlation between GEBV and true breeding values of all offspring individuals. Bias of GEBV predictions was computed as the regression coefficient of true breeding values on GEBV. Coancestry coefficients were computed as the means of realized genomic relationships between top GEBV rankings of 20 males and 200 females.

Pairwise genomic relationships between animals that were half-sibs, full-sibs, paternal half-sibs or genomic-close full-sibs were used to investigate the distribution of relationships. Half-sibs were offspring from the same sires, but different dams based on pedigree. Paternal half-sibs were offspring from the same sire, that could, but not necessarily, be from the same dam. Genomic-close full-sibs were full-sibs that became members of the same group after applying the unsupervised clustering based on genotyping or supervised clustering based on genomic relationships. All pairwise genomic relationships of group members from all 100 replicates were combined and used to calculate means and standard deviations.

6.3. Results

Realized genomic relationships in breeding scheme were calculated for half-sibs, paternal half-sibs, full-sibs and groups of genomic-close full-sibs that were grouped by either unsupervised clustering based on genotypes or supervised clustering based on genomic relationships (Table 6.2). As expected, means of realized genomic relationships were roughly 0.50 for full-sibs and 0.25 for half-sibs (Figure 6.1). Paternal half-sib relationships were mixture of full-sib and half-sib relationships. Mean of relationships between genomic-close full-sibs grouped by the supervised clustering method was the highest at 0.55 followed by genomic-close full-sibs grouped by the unsupervised clustering method at 0.54. Relationships between genomic-close full-sibs pooled by grouping methods that were based on genomic information had lower standard deviation than the relationships between full-sibs.

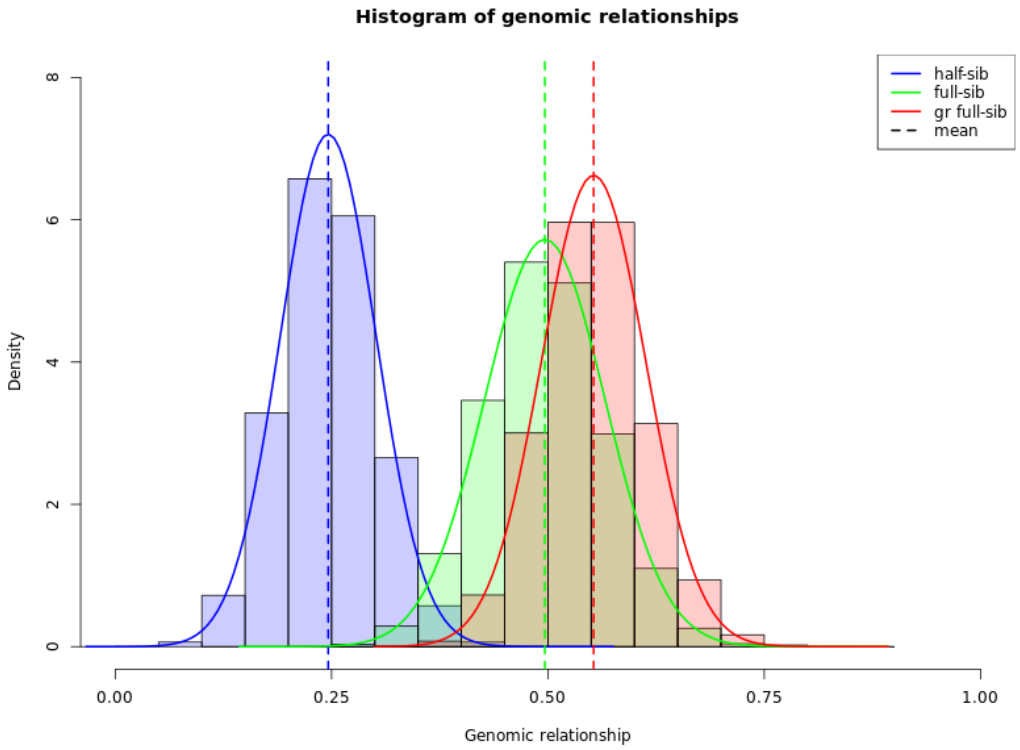


Figure 6.1. Distribution of realized genomic relationship between half-sibs (blue line and light-blue bars), full-sibs (green line and light-green bars) and genomic-close full-sibs (full-sibs grouped by the supervised clustering method based on genomic relationship (red line and pink bars)). Broken vertical lines were means of the genomic relationships.

Table 6.2: Means and standard deviations of realized genomic relationships between half-sibs, paternal half-sibs, full-sibs and genomic-close full-sibs that were grouped by unsupervised clustering based on genotypes or by supervised clustering based on genomic relationships.

Relationships	Mean	Standard deviation
Half-sibs	0.246	0.055
Paternal half-sibs	0.270	0.093
Full-sibs	0.496	0.070
Genomic-close full-sibs grouped by supervised clustering based on genomic relationships	0.553	0.060
Genomic-close full-sibs grouped by unsupervised clustering based on genotypes	0.538	0.067

Variance components estimated from group records were consistent with those estimated from individual records (Table 6.3), and estimates were not significantly different from simulated values. However,

variance components estimated from group records had a higher standard deviation than those estimated from individual records.

Table 6.3: Estimates of additive genetic variance (σ^2_a) and residual variance (σ^2_e) (mean over 100 replicates ± standard deviation) estimated from individual records and from group records

Records	Model	σ^2_a (SD)	σ^2_e (SD)
Simulated parameters		0.30	0.70
Individual records	$GBLUP_i$	0.300 (0.030)	0.698 (0.022)
Group records from supervised clustering method based on genomic relationships	$GBLUP_{gr}$	0.302 (0.042)	0.691 (0.048)
Group records from unsupervised clustering method based on genotypes	$GBLUP_{gr}$	0.301 (0.043)	0.693 (0.050)
Group records from random grouping of full-sibs	$GBLUP_{gr}$	0.298 (0.045)	0.695 (0.052)
Group records from random grouping of paternal half-sibs	$GBLUP_{gr}$	0.301 (0.062)	0.695 (0.050)

Note: Models $GBLUP_i$ and $GBLUP_{gr}$ are GBLUP model for individual records and group records, respectively. SD is standard deviations over 100 replicates.

Accuracy and bias of GEBV of different prediction models from individual and group records are presented in Table 6.4. As expected, accuracy of GEBV was higher from individual records than from group records. When group records were used to predict GEBV, accuracy of GEBV depended on grouping methods. Accuracies of GEBV became smaller when the realized genomic relationships between group members reduced. Grouping methods based on genomic information led to higher accuracy of GEBV than the random grouping methods based on pedigree information. Group records from supervised clustering based on genomic relationship led to the highest accuracy and the lowest standard deviation of the accuracy compared to group records from other grouping methods. Group records from random grouping of paternal half-sibs resulted in the lowest accuracy of GEBV prediction and the highest standard deviation of the accuracy.

Table 6.4: Accuracy of GEBV, bias of prediction and coancestry coefficients of top ranking animals (mean over 100 replicates ± standard deviation) on GEBV estimated from individual records and from group records

Records	Model	Accuracy (SD)	Bias (SD)	Coancestry coefficients (SD)
Individual records	$GBLUP_i$	0.825 (0.020)	1.011 (0.041)	0.036 (0.009)
Group records from supervised clustering method based on genomic relationships	$GBLUP_{gr}$	0.762 (0.028)	1.007 (0.054)	0.041 (0.010)
Group records from unsupervised clustering method based on genotypes	$GBLUP_{gr}$	0.758 (0.030)	1.009 (0.054)	0.041 (0.009)
Group records from random grouping of full-sibs	$GBLUP_{gr}$	0.749 (0.032)	1.015 (0.060)	0.043 (0.010)
Group records from random grouping of paternal half-sibs	$GBLUP_{gr}$	0.682 (0.040)	1.017 (0.092)	0.049 (0.010)

Coancestry coefficients were computed for top GEBV ranking 20 males and 200 females with GEBV estimated from individual records and from group records (Table 6.4). The use of individual records led to lower coancestry coefficients between top GEBV ranking animals than group records. Grouping methods based on genomic information led to lower coancestry coefficients than the random grouping methods based on pedigree information.

Sensitivity analysis

In SS1-4, variance components estimated from group records were consistent with those estimated from individual records. Bias of GEBV estimated from individual records and group records did not show a clear difference, and the values were close to 1. However, standard deviations of variance estimates and bias of GEBV over 100 replicates were higher for group records than for individual records. Results on variance estimates and bias of GEBV of SS1-4 were not shown. Genomic relationships between group members, accuracy of GEBV and coancestry coefficients of top ranking animals were comparable measures between scenarios in SS1-4. Just like in the main study, the supervised clustering method based on genomic relationships generally led to a higher genomic relationships between group members, higher accuracy of GEBV and lower coancestry coefficients of top ranking animals than random grouping of full-sibs and random grouping of paternal half-sibs.

In SS1, group size and number of groups per full-sib family were the same as in the main study, but family sizes were varied, and therefore after grouping procedures, there were surplus offspring that did not

belong to any groups or get phenotypes. With 0 (main study), 16 (SS1) and 32 (SS1) surplus offspring, relationships between genomic-close full-sibs that were grouped by supervised clustering were 0.55, 0.60 and 0.62, respectively (Table 6.2 & 6.5). The increase in number of surplus offspring tended to increase accuracy of GEBV estimated from group records of the genomic-close full-sibs (Table 6.4 & 6.5). Meanwhile, the change in number of surplus offspring did not affect relationships between group members or accuracy of GEBV for scenarios with groups that were formed by random grouping of full-sibs or random grouping of paternal half-sibs. The relative increase in accuracy of GEBV from the use of random grouping of full-sibs to the use of supervised clustering based on genomic information was 3.9% and 4.5% when number of surplus offspring were 16 and 32 per full-sib family, respectively.

Table 6.5: Comparable measures between scenarios for sensitivity simulation 1 when group size was constant at 4 animals per group, number of groups were constant at 4 groups per full-sib family, and number of surplus genotyped offspring without phenotypes were varied at 16 and 32 per full-sib family.

Variables	Individual records	Group records		
		Supervised clustering method	Random grouping of full-sibs	Random grouping of paternal half-sibs
Surplus offspring: 16 per full-sib family				
Genomic relationships (SD)		0.602 (0.053)	0.496 (0.070)	0.270 (0.093)
Accuracy (SD)	0.824 (0.020)	0.773 (0.028)	0.744 (0.032)	0.678 (0.040)
Coancestry coefficients (SD)	0.036 (0.008)	0.042 (0.010)	0.043 (0.009)	0.049 (0.011)
Surplus offspring: 32 per full-sib family				
Genomic relationships (SD)		0.622 (0.052)	0.497 (0.070)	0.271 (0.094)
Accuracy (SD)	0.822 (0.021)	0.776 (0.028)	0.743 (0.032)	0.677 (0.040)
Coancestry coefficients (SD)	0.035 (0.007)	0.042 (0.008)	0.043 (0.009)	0.047 (0.010)

In SS2, family size was kept constant, and it was the same as family size in the main study, but number of groups per full-sib family, and thus group sizes, were varied. With group records formed from the supervised clustering method, random grouping of full-sibs and random grouping of paternal half-sibs, increasing number of groups per full-sib family all led to an increase in accuracy of GEBV and a decrease in coancestry coefficients of top ranking animals (Table 6.6). With group records formed from the supervised clustering method, changing number of groups per full-sib family to 2 or 8 changed the genomic relationships between group members to 0.522 and 0.589, respectively. However, the benefit in term of accuracy of GEBV due to the use of supervised clustering grouping method did not show a clear difference between different number of groups because the relative differences in accuracy of GEBV

between methods of supervised clustering and random grouping of full-sibs, for example, were 1.01, 1.02 and 1.01 for scenarios with 2, 4 and 8 groups per full-sib family, respectively.

Table 6.6: Comparable measures between scenarios for sensitivity simulation 2 when family size was constant at 16 offspring per full-sib family, group sizes were varied at 8 and 2 animals per group, and number of groups were varied at 2 and 8 groups per full-sib family.

Measures	Individual records	Group records		
		Supervised clustering method	Random grouping of full-sibs	Random grouping of paternal half-sibs
Number of groups: 2 groups per full-sib family				
Genomic relationships (SD)		0.589 (0.053)	0.496 (0.070)	0.270 (0.093)
Accuracy (SD)	0.825 (0.020)	0.794 (0.025)	0.783 (0.026)	0.755 (0.030)
Coancestry coefficients (SD)	0.036 (0.009)	0.038 (0.009)	0.039 (0.009)	0.041 (0.009)
Number of groups: 8 groups per full-sib family				
Genomic relationships (SD)		0.522 (0.066)	0.496 (0.070)	0.270 (0.093)
Accuracy (SD)	0.825 (0.020)	0.736 (0.033)	0.726 (0.035)	0.620 (0.051)
Coancestry coefficients (SD)	0.036 (0.009)	0.044 (0.010)	0.045 (0.010)	0.059 (0.012)

In SS3, number of groups were kept constant at 4 groups per family, but family sizes, and therefore group sizes, were varied. With group records formed from different grouping methods in SS3, increasing family sizes led to an increase in accuracy of GEBV and coancestry coefficients of top ranking animals (Table 6.7). With group records formed from the supervised clustering method, increasing family sizes led to a decrease in genomic relationships between group members. The relative differences in accuracy of GEBV between methods of supervised clustering and random grouping of full-sibs, for example, were 1.01, 1.02, 1.02 and 1.03 for scenarios with family sizes of 8, 16, 32 and 48 offspring per full-sib family, respectively.

Table 6.7: Comparable measures between scenarios for sensitivity simulation 3 when number of groups were constant at 4 groups per full-sib family, group sizes were varied at 2, 8 and 12 animals per group, and family sizes were varied at 8, 32 and 48 offspring per full-sib family.

Variables	Individual records	Group records		
		Supervised clustering method	Random grouping of full-sibs	Random grouping of paternal half-sibs
Family size of 8				
Genomic relationships (SD)		0.565 (0.057)	0.497 (0.070)	0.268 (0.091)
Accuracy (SD)	0.770 (0.028)	0.734 (0.034)	0.726 (0.035)	0.692 (0.041)
Coancestry coefficients (SD)	0.024 (0.004)	0.026 (0.004)	0.027 (0.004)	0.028 (0.005)
Family size of 32				
Genomic relationships (SD)		0.543 (0.063)	0.497 (0.070)	0.271 (0.093)
Accuracy (SD)	0.870 (0.014)	0.785 (0.026)	0.766 (0.028)	0.674 (0.040)
Coancestry coefficients (SD)	0.049 (0.016)	0.060 (0.019)	0.063 (0.020)	0.078 (0.026)
Family size of 48				
Genomic relationships (SD)		0.539 (0.064)	0.497 (0.070)	0.271 (0.094)
Accuracy (SD)	0.893 (0.011)	0.799 (0.023)	0.774 (0.025)	0.671 (0.041)
Coancestry coefficients (SD)	0.053 (0.017)	0.068 (0.021)	0.072 (0.021)	0.091 (0.032)

Breeding structures, group sizes and family sizes of SS4 were the same as those of the main study, but the genome structure was different, in which the genome size of SS4 was longer (3000cM *versus* 916cM). Accuracy of GEBV from individual records and group records was higher in the main study than SS4 (Table 6.8). Means of genomic relationships between full-sibs or between paternal half-sibs were similar for the main study and SS4, but standard deviations of the relationships were lower for SS4. The genomic relationship between genomic-close full-sibs that were grouped by the supervised clustering method was higher for the main study than for SS4. The relative differences in accuracy of GEBV between methods of supervised clustering and random grouping of full-sibs, for example, were 1.02 and 1.01 for scenarios in the main study and SS4, respectively.

Table 6.8: Comparable measures for sensitivity simulation 4 when the genome of 30 chromosomes with total length of 3000cM was used.

Variables	Individual records	Group records		
		Supervised clustering method	Random grouping of full-sibs	Random grouping of paternal half-sibs
Genomic relationships (SD)		0.532 (0.036)	0.496 (0.042)	0.269 (0.080)
Accuracy (SD)	0.729 (0.030)	0.671 (0.039)	0.662 (0.040)	0.583 (0.055)
Coancestry coefficients (SD)	0.036 (0.009)	0.042 (0.010)	0.044 (0.010)	0.054 (0.011)

6.4. Discussion

With GBLUP models, variance components estimated from group records and individual records were consistent. However, the standard errors of estimates from group records were larger than those from individual records. Similar conclusions on using group records to estimate variance components were also drawn in other studies that used pedigree-based BLUP models (Biscarini *et al.*, 2008; Olson *et al.*, 2006; Su *et al.*, 2018). Compared to these studies (Biscarini *et al.*, 2008; Olson *et al.*, 2006; Su *et al.*, 2018), the main modification of our model for group records was the use of realized genomic relationship matrix instead of numerator genetic relationship matrix. With numerator genetic relationship matrix, full-sibs in the same group have equal EBV whereas with genomic relationship matrix, full-sibs in the same group can have different EBV. The benefit of genomic information over pedigree information, in term of accuracy of prediction, has been well documented in simulations (Hayes *et al.*, 2009; Meuwissen *et al.*, 2001; Putz *et al.*, 2018) and empirical studies of chicken (Alemu *et al.*, 2016; Momen *et al.*, 2017; Wolc *et al.*, 2011), cattle (Gao *et al.*, 2018; Lourenco *et al.*, 2015a) and pig (Christensen *et al.*, 2012; Guo *et al.*, 2015; Putz *et al.*, 2018) breeding schemes for individual records. The increase in accuracy of GBLUP prediction from individual records is attributable to better measuring the relationships between animals and a better prediction of the Mendelian sampling terms (Hayes *et al.*, 2009). These attributes of genomic information should also apply to GBLUP models for group records, thus increase accuracy of prediction compared to the use of pedigree-based models for group records.

For the same number of individuals, accuracy of GEBV based on group records was lower than those based on individual records. Coancestry coefficients of selected animals based on group records were also higher than those based on individual records. While results from group records cannot compete with results from individual records, the number of phenotypes to be recorded are also are unequal between group and individual record data. At commercial production environment level, group records are sometimes the only available phenotypes.

When group records were analyzed, accuracy of estimates for GEBV depended on relationships between group members. The accuracy increased when group members were more close-related. This also has been shown in Olson *et al.* (2006), Peeters *et al.* (2013) and Su *et al.* (2018). Allocation of animals based on sires resulted in higher accuracy of EBV than the allocation based on maternal grand sire (Olson *et al.*, 2006). Prediction of EBV and variance components was more accurate with group records of animals from the same family than group records of animals from two different families (Peeters *et al.*, 2013; Su *et al.*, 2018). A possible reason could be that more of the phenotypic variance at the group level is explained by

additive genetic (co)variance when increasing the level of relationships between individuals within a group (Su *et al.*, 2018).

In addition to an increase in accuracy of GEBV through the use of the genomic relationship matrix in a GBLUP model, genomic information gave additional accuracy of GEBV through optimized grouping. Our proposed grouping methods based on genomic information resulted in higher relationship coefficients between individuals within groups than random grouping based on pedigree. The higher relationship coefficients from the proposed grouping methods led to a higher accuracy of GEBV when group records were used. Compared to random grouping of full-sibs, the improvement in accuracy was 1.2-1.7% with genomic information. However, while higher accuracies are preferred, the grouping methods based on genomic information require individual genotyping before animals are transferred to phenotype testing facilities. The small improvement in accuracy of GEBV may not offset the genotyping cost in a situation where full-sib groups can be made without obtaining genomic information. For situations where only paternal half-sib groups could be produced and full-sibs could not be identified, the advantage would be 11.1-11.7%. Our approach could be used when the aim is to obtain feed efficiency records for a commercial testing environment or egg production from village household chickens e.g. African Chicken genetic gains program (ACCG, 2014). Another application lies with genomic selection in fish breeding programs where mating and reproduction is natural and sib information is absent (Joshi *et al.*, 2018). When genotyping information is available prior to group testing, grouping based on genomic information could give additional "rewards" in the form of accuracy to genomic selection in such breeding programs.

Coancestry coefficients were also reduced with grouping methods based on genomic information compared to random grouping based on pedigree. In our study, the coancestry coefficients were defined as realized genomic relationships between the top GEBV ranking 20 males and 200 females. Therefore, the coancestry coefficients are indications of future inbreeding when GEBV estimated from group records are used for selection. The use of more close-related animals to form groups can have two opposite consequences for the coancestry coefficients. One consequence leads to an increase in coancestry coefficients. As the more close-related animals in the same group have the same phenotypic group records, GEBV between those animals are more similar, thus increasing co-selection. The other consequence leads to a reduction in coancestry coefficients because the use of more close-related animals to form groups increases accuracy of GEBV prediction from group records, thus reducing co-selection. The latter benefit is only obtained with the GBLUP model because a reduction in co-selection due to increasing accuracy of prediction does not occur with pedigree-based BLUP. The EBV predicted from group records with pedigree-based BLUP are identical for full-sibs in the same group. With the prediction of GBLUP

model from group records, full-sibs in the same group can have different GEBV, thus selected animals with top GEBV ranking can come from different groups and different families. The effect of increasing accuracy of prediction was more pronounced when the more close-related animals were used to form groups. Therefore, compared to random grouping of full or half-sibs based on pedigree, a reduction in coancestry coefficients of selected candidates was observed with grouping methods based on genomic information. Of the two proposed grouping methods based on genotyping information, supervised clustering based on genomic relationship had higher accuracy, was less computationally demanding and is more applicable in practice than unsupervised clustering based on genotyping. Unsupervised clustering analysis with the STRUCTURE program uses genotyping data to infer population structure and assign individuals to clusters, each of which is characterized by a set of allele frequencies at each locus (Pritchard *et al.*, 2000). Updating criteria for inferring population structure and assigning individuals are similarity or homogeneity of alleles between individuals in clusters and Hardy-Weinberg equilibrium of alleles in clusters. With this inferred population structure, half-sibs from each sire were assigned to one group when unsupervised clustering analysis of STRUCTURE program was applied to all offspring. Full-sibs from each family would be assigned to one group when the unsupervised clustering analysis was applied to paternal half-sibs from each sire. Therefore, unsupervised clustering analysis of STRUCTURE was applied to each of full-sib families. After that, membership coefficients of individuals that belong to clusters had to be used to arrive at equally sized groups. Because of this re-arrangement of animals between groups, the advantage of the unsupervised clustering method to pool animals with genomic similarity into groups was reduced. Unsupervised clustering based on genotypes is not ideal for assigning animals to groups when testing facilities often have fixed capacity for group sizes and number of groups. In addition, the unsupervised clustering analysis of STRUCTURE for each full-sib family is hundreds of times more computation-expensive than the grouping method of supervised clustering based on genomic relationships. In contrast, our proposed grouping method of supervised clustering assigns individuals to groups based on genomic relationship matrix that was calculated from genotyping data. This grouping method uses a relatively simple evolutionary algorithm to cluster animals into predefined number of groups and desired group sizes.

Supervised clustering based on genomic relationships was carried out for each full-sib family because the probability of allocating half-sibs into the same group was very unlikely. The overlap of the distribution of full-sib and half-sib relationships is very small (Figure 6.1). When supervised clustering was applied to form groups from the whole population at once, members within a group were always from the same full-sib family. Grouping from the whole population was time-consuming, thus only few replicates were tested

(results not shown). However, it is good to realize that when family relationships are not available to apply grouping within full-sib families, the same benefits of grouping based on genomic relationships can be obtained with additional computational effort. The same principles for animal grouping based on genomic relationships can be also applied to paternal half-sibs, half-sibs and all testing candidates when number of full-sibs per family are smaller than the intended group sizes. Compared to grouping based on pedigree information, grouping based on genomic relationships does not lead to an increase in genomic relationships between animals within groups in one specific situation; only when number of full-sibs per family are equal to group sizes. In that situation grouping based on pedigree and grouping based on genomic relationships will give the same result.

Benefits of supervised clustering based on genomic information were further elaborated when there were surplus offspring that were available for grouping procedures, but at the end some were not assigned to any groups for phenotyping. Compared to random grouping based on pedigree information, surplus offspring available for supervised clustering based on genomic information increased genomic relationships between group members, and improved accuracy of GEBV estimated from group records. The improvement in accuracy of GEBV was up to 4.5% from the use of random grouping of full-sibs or 14.7% from the use of random grouping of paternal half-sibs to the use of supervised clustering based on genomic relationships. Other factors that affected genomic relationships between group members formed from the supervised clustering method were family sizes, number of groups, group sizes and genome structure. However, in term of accuracy of GEBV, benefits of supervised clustering based on genomic information depended little on those factors.

Genomic relationships between group members that are formed from the supervised clustering method depended on genome structure that may be related to size of the genome in our simulation. Increasing genomic size reduces genomic relationship between group members because standard deviation of genomic relationship between full-sibs and between half-sibs reduces. For example, the standard deviation of genomic relationships between full-sibs is approximately equal to $\frac{0.5}{(2n_l)^{0.5}}$, where n_l is number of independent loci of the genome (VanRaden, 2007). The standard deviation would be zero if number of loci is enormous. However, the standard deviation does not fall below about 0.035 because the loci are usually linked rather than independent (VanRaden, 2007). In addition, increasing genome size reduces accuracy of GEBV from GBLUP models as shown in formula by Daetwyler *et al.* (2010) that the accuracy is dependent on the genome length in Morgans. The simulated genome structure in the main study was constructed based on the genome of chicken.

6.5. Conclusions

Two grouping methods based on genomic information were proposed to improve accuracy of prediction from group records. Variance components and GEBV from group records were estimated using GBLUP models. It was found that estimates of variance components from group records were consistent with those from individual records and with their true values. Our two proposed grouping methods based on genomic information led to higher genomic relationship between group members, and prediction from group records pooled by these two methods had higher accuracy of GEBV prediction compared to random grouping based on pedigree information. In addition, grouping based on genomic information led to lower coancestry coefficients of selected candidates than random grouping of paternal half-sibs and random grouping of full-sibs. Of the two proposed methods, supervised clustering based on genomic relationships was superior in term of computation requirements, applicability and accuracy of GEBV prediction. Benefits of supervised clustering based on genomic information were further elaborated for accuracy of GEBV estimated from group records when there were surplus offspring that were available for grouping procedures. Genomic relationship between group members that were formed from the supervised clustering method depended on factors of family sizes, number of groups, group sizes and genome structure, but these factors had little influence on the benefits of the grouping method in term of accuracy of GEBV. In summary, genotyping information can be utilized to increase accuracy of prediction from group records in two ways: genomic-based prediction and optimized grouping.

References:

ACCG 2014, 'African Chicken Genetic Gains', accessed: 02/04/2019, africacgg.net.

Alemu, SW, Calus, MPL, Muir, WM, Peeters, K, Vereijken, A & Bijma, P 2016, 'Genomic prediction of survival time in a population of brown laying hens showing cannibalistic behavior', *Genetics Selection Evolution,* vol. 48, no. 1, pp. 68.

Biscarini, F, Bovenhuis, H & van Arendonk, JA 2008, 'Estimation of variance components and prediction of breeding values using pooled data', *Journal of Animal Science,* vol. 86, no. 11, pp. 2845-52.

Calus, MPL, Mulder, HA & Bastiaansen, JWM 2011, 'Identification of Mendelian inconsistencies between SNP and pedigree information of sibs', *Genetics Selection Evolution,* vol. 43, no. 1, pp. 34.

Christensen, OF, Madsen, P, Nielsen, B, Ostersen, T & Su, G 2012, 'Single-step methods for genomic evaluation in pigs', *Animal,* vol. 6, no. 10, pp. 1565-71.

Chu, TT, Alemu, SW, Norberg, E, Sørensen, AC, Henshall, J, Hawken, R & Jensen, J 2018, 'Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 52.

Daetwyler, HD, Pong-Wong, R, Villanueva, B & Woolliams, JA 2010, 'The impact of genetic architecture on genome-wide evaluation methods', *Genetics,* vol. 185, no. 3, pp. 1021-1031.

Falush, D, Stephens, M & Pritchard, JK 2003, 'Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies', *Genetics,* vol. 164, no. 4, pp. 1567-87.

Gao, H, Koivula, M, Jensen, J, Strandén, I, Madsen, P, Pitkänen, T, Aamand, GP & Mäntysaari, EA 2018, 'Short communication: Genomic prediction using different single-step methods in the Finnish red dairy cattle population', *Journal of Dairy Science,* vol. 101, no. 11, pp. 10082-10088.

Guo, X, Christensen, OF, Ostersen, T, Wang, Y, Lund, MS & Su, G 2015, 'Improving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method', *J Anim Sci,* vol. 93, no.

Hayes, BJ, Visscher, PM & Goddard, ME 2009, 'Increased accuracy of artificial selection by using the realized relationship matrix', *Genet Res,* vol. 91, no.

Joshi, R, Woolliams, JA, Meuwissen, T & Gjoen, HM 2018, 'Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits', *Heredity (Edinb),* vol. 120, no. 5, pp. 452-462.

Lourenco, DA, Tsuruta, S, Fragomeni, BO, Masuda, Y, Aguilar, I, Legarra, A, Bertrand, JK, Amen, TS, Wang, L, Moser, DW & Misztal, I 2015a, 'Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus', *J Anim Sci,* vol. 93, no. 6, pp. 2653-62.

Lourenco, DAL, Fragomeni, BO, Tsuruta, S, Aguilar, I, Zumbach, B, Hawken, RJ, Legarra, A & Misztal, I 2015b, 'Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken', *Genetics Selection Evolution,* vol. 47, no. 1, pp. 56.

Madsen, P & Jensen, J 2013, 'DMU: a user's guide'. "*A package for analysing multivariate mixed models, Version 6, release 5.2. Home page: http://dmu.agrsci.dk/*",

Meuwissen, THE, Hayes, BJ & Goddard, ME 2001, 'Prediction of total genetic value using genome-wide dense marker maps', *Genetics,* vol. 157, no. 4, pp. 1819-29.

Momen, M, Mehrgardi, AA, Sheikhy, A, Esmailizadeh, A, Fozi, MA, Kranis, A, Valente, BD, Rosa, GJM & Gianola, D 2017, 'A predictive assessment of genetic correlations between traits in chickens using markers', *Genetics Selection Evolution,* vol. 49, no. 1, pp. 16.

Olson, KM, Garrick, DJ & Enns, RM 2006, 'Predicting breeding values and accuracies from group in comparison to individual observations', *Journal of Animal Science,* vol. 84, no. 1, pp. 88-92.

Pedersen, L, Sørensen, A, Henryon, M, Ansari-Mahyari, S & Berg, P 2009, 'ADAM: A computer program to simulate selective breeding schemes for animals', *Livestock Science,* vol. 121, no. 2, pp. 343-344.

Peeters, K, Ellen, ED & Bijma, P 2013, 'Using pooled data to estimate variance components and breeding values for traits affected by social interactions', *Genetics Selection Evolution,* vol. 45, no.

Pritchard, JK, Stephens, M & Donnelly, P 2000, 'Inference of population structure using multilocus genotype data', *Genetics,* vol. 155, no. 2, pp. 945-59.

Putz, AM, Tiezzi, F, Maltecca, C, Gray, KA & Knauer, MT 2018, 'A comparison of accuracy validation methods for genomic and pedigree-based predictions of swine litter size traits using Large White and simulated data', *J Anim Breed Genet,* vol. 135, no. 1, pp. 5-13.

R Core Team 2018, 'R: A language and environment for statistical computing'. "*R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.*",

Su, G, Madsen, P, Nielsen, B, Ostersen, T, Shirali, M, Jensen, J & Christensen, OF 2018, 'Estimation of variance components and prediction of breeding values based on group records from varying group sizes', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 42.

VanRaden, P 2007, 'Genomic measures of relationship and inbreeding', *Interbull Bull*, no. 37, pp. 33-36.

VanRaden, PM 2008, 'Efficient methods to compute genomic predictions', *J Dairy Sci,* vol. 91, no.

Wolc, A, Stricker, C, Arango, J, Settar, P, Fulton, JE, O'Sullivan, NP, Preisinger, R, Habier, D, Fernando, R, Garrick, DJ, Lamont, SJ & Dekkers, JCM 2011, 'Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model', *Genetics Selection Evolution,* vol. 43, no. 1, pp. 5.

7

General Discussion

The study presented in Chapter 2 and 3 investigated GxE interactions due to environmental differences for purebred broiler chicken. Environmental differences between B and C led to significant re-ranking of EBV of birds. This Chapter 7 will extend the discussions to GxE interactions for crossbred chicken raised in multiple commercial production environments. GxE interactions due to micro-environmental differences will also be discussed. Chapter 4 showed the development and comparison of statistical models for improving accuracy of prediction of BW traits in broilers, and this chapter will discuss other available tools for the development and comparison of statistical models. The tools are goodness-of-fit of the models, accuracy of EBVs and cross-validation approaches. Different prediction models for group records in Chapter 5 and 6 will be compared to the model for individual records and the model that includes the group effect and permanent environmental maternal effect. Theoretical benefits of genomic information for breeding programs affected by GxE will be discussed.

7.1. GxE interactions in breeding programs

In Chapter 1 (General Introduction), the concept and definition of GxE interactions due to differences between two environments (B and C) were presented. However, in practice, the context of GxE interactions in the breeding programs for poultry is much more complex. For example, it can involve multiple C environments and the use of cross-bred production birds. This section extends the discussions on the GxE interactions in these situations.

Selection on crossbred or purebred performances in production environments

Chapter 2-6, which investigates of GxE and solutions to improve accuracy of selection for traits in C, focused on the use of purebred performances in C environments. However, in poultry and pig industry, production animals are 3- or 4-way crossbreds. The use of crossbreeding will benefit from heterosis, breed complementarity and flexibility of creating products for different markets (Wientjes and Calus, 2017). Therefore, the ultimate breeding goal of breeding programs for poultry and pigs should be to increase performances of crossbred animals in C environments. In this section, I will discuss about advantages and disadvantages of selection based on crossbred performances in C environments and purebred performances in B environments.

The use of performances of purebred in B and crossbred in C for selection has been explored in breeding programs for pigs (Godinho *et al.*, 2018; Habier *et al.*, 2007; Hidalgo *et al.*, 2015; Nielsen *et al.*, 2016; Xiang *et al.*, 2016) and for laying egg chicken (Wei and Werf, 1995). Similar to modelling GxE interactions of a trait measured in B and C environments, a trait measured on purebred animals in B and crossbred animals in C can be treated as two correlated traits. Wientjes and Calus (2017) summarized genetic parameter

estimates of traits measured on purebred and crossbred pigs from 27 studies published between 1964 and 2017. The authors shows that on average, genetic correlations between the traits measured on purebred and crossbred animals were 0.66, 0.69, 0.67, 0.54 and 0.67 for growth, meat amount, meat quality, fertility, and feed efficiency traits, respectively. The genetic correlation between the traits measured on purebred and crossbred animals is a combination of three components: genotype by genotype interactions due to the differences in genetic background of purebred versus crossbred animals, genotype by environment interactions due to the environmental differences between rearing conditions of purebred and crossbred animals, and differences in the definitions or measurements between traits recorded on purebred and crossbred animals (Wientjes and Calus, 2017). Each of these three components can reduce the genetic correlation between the traits measured on purebred and crossbred animals. Therefore, it is more logical to carry out selection based on performances of crossbred animals in C than performances of purebred animals in C because genetic correlations between traits measured on purebred animals in B and C accounts for only the environmental differences between selection candidates and production animals.

However, the additive genetic relationship coefficients between selection candidates and crossbred production animals are often low due to one or more generation of multiplication of selected genotypes, thus accuracy of selecting purebred animals for crossbred performances can be low. Crossbred animals in C can be half-sibs, progeny or distant descendants of purebred selection candidates in B (Wientjes and Calus, 2017). On the contrary, the relationships between selection candidates in B and purebred birds in C can be full-sibs and half-sibs (Chu *et al.*, 2018; Kapell *et al.*, 2012). To my knowledge, selection based on performances of crossbred animals in C environments were only carried out in breeding programs for pigs, but not for broilers. This may be related to complexity of recording pedigree information of crossbred broilers in C. In addition, compared to selection based on purebred birds in C, the progress of genetic gain can be slower due to a longer generation interval of selection based on crossbred records from C. Also, GxE interactions due to environmental differences are relatively lower in pig than in poultry. When the genetic correlation between the traits measured on purebred and crossbred pigs are decomposed into three components, Wientjes and Calus (2017) shows that genotype by genotype interactions contribute large effects on the genetic correlation while GxE interactions contribute relatively small effects on the genetic correlations. In contrast, GxE interactions due to the environmental differences are substantial for growth traits of broilers with genetic correlations between traits measured in the breeding and production environments in range 0.48-0.54 (Chapter 2) and 0.46-0.69 (Kapell *et al.*, 2012). However, genotype by genotype interactions are relatively low for chicken. Duenk *et al.* (2019) shows that the

genetic correlation between purebred and 3-way crossbred performances of broilers measured in the same environment was high, ranging from 0.90-0.96 for BW at 5 weeks of age. The high correlation implies low genotype by genotype interactions for broiler BW5.

Selection for multiple production environments

In Chapter 2, 3 and 5, two distinct classes of environments are defined as B and C. Conditions of production environments in GxE breeding programs are replicated by resembling broad commercial-like conditions (Kapell *et al.*, 2012) (Chapter 2). However, conditions of the commercial production environments often vary from farm to farm for hygienic conditions (different levels), diseases (types of specific diseases and level of pathogen burden), diets (type of feed, protein levels and feeding regimes), management (management of litter, light and temperature) and stocking density. When production animals are raised in a broad range of environmental conditions, selection for robustness or resilience of animals might be more appropriate.

To model GxE interactions in case of multiple environments, multi-trait model and reaction norm model are commonly used. The models described below are pedigree-based BLUP models. The multi-trait model treats a trait measured in different environments as different traits:

$y_{ijk} = \mu + E_j + a_{ij} + e_{ijk}$ (7.1)

where y_{ijk} is phenotypic record *k* of genotype *i* in environment *j*; μ is the mean; *Ej* is environment *j*; a_{ij} is the additive genetic effect of genotype *i* for performance in environment *j*; e_{ijk} is the residual. The additive genetic effects in different environments are assumed to be normally distributed: N[**0**, $\mathbf{A} \otimes \mathbf{V_a}$], where **A** is a relationship matrix, $\mathbf{V_a}$ is a *n*x*n* covariance matrix, and *n* is number of environments. The residual environmental terms are also normally distributed: N[**0**, $\mathbf{I} \otimes \mathbf{V_e}$], where **I** is identity matrix for individuals, and $\mathbf{V_e}$ is a *n*x*n* covariance matrix with off-diagonal elements equal to zero, as each animal can have records in one environment due to restriction of move animals between environment.

The reaction norm model as used in Sartori *et al.* (2018) that uses Legendre polynomials on level of environmental challenge can be:

$$y_{ijk} = \mu + \sum_{l=0}^{nf} \phi_{jl}\beta_l + \sum_{l=0}^{nr} \phi_{jl}a_{il} + e_{ijk} \quad (7.2)$$

where y_{ijk} is phenotypic record *k* of genotype *i* in environment *j*; μ is mean; ϕ_{jl} is the *l*th Legendre polynomial for phenotypic record on environmental parameter *j*; β_l is the *l*th fixed regression coefficients; a_{il} is the *l*th random regression for additive genetic effects; *nf* and *nr* are the order of polynomials for fixed regression and additive genetic effects; e_{ijk} is the residual. In case of linear reaction norm model (*nr* = 1), the additive genetic effects for the intercept and slope are assumed to follow a bivariate normal

distribution: N[0, $\mathbf{A} \otimes \mathbf{V}_{\mathbf{aRN}}$], with $\boldsymbol{V}_{\boldsymbol{a_{RN}}} = \begin{bmatrix} \sigma^2_{a_{int}} & \sigma_{a_{int},a_{lin}} \\ \sigma_{a_{int},a_{lin}} & \sigma^2_{a_{lin}} \end{bmatrix}$, where $\sigma^2_{a_{int}}$ and $\sigma^2_{a_{lin}}$ are additive genetic variances for the intercept and slope, and $\sigma_{a_{int},a_{lin}}$is additive genetic covariance between the intercept and slope.

GxE interactions are present when correlations between traits measured in environments are lower than 1 for the multivariate model (7.1), or additive genetic variances for *l*th random regression (*l* ≥1) such as $\sigma^2_{a_{lin}}$ are larger than 0 for the reaction norm model (7.2). The multivariate model would have problems with estimation of variance components and model convergence when number of environments are large. The model used in Chapter 2, 3 and 5 is similar to the multi-trait model (7.1), except that only two distinct classes of environments including B and C environments were defined. The reaction norm model describes the expected reaction of genotypes to a specific environmental parameter or level of environmental challenge. The reaction norm model have fewer parameters than the multivariate model when there are more than two environments. In comparisons between the two models, the multivariate model (7.1) is better than the reaction norm model (7.2) in term of estimation of GxE, flexibility of variance-covariance structure and genetic interpretation of GxE (Mulder, 2007). However, the reaction norm model is better in term of predictability of phenotypes, biological interpretation of GxE and selection for robust animals (Mulder, 2007).

The reaction norm model describes the change in the genetic effect per unit change in an environmental parameter such as temperatures, level of pathogen burden, level of protein or combination of these effects. For example, the reaction norm models were used to explore genetic variation of growth traits as a function of temperatures for pigs (Zumbach *et al.*, 2008), beef cattle (Santana *et al.*, 2016) and rainbow trout (Janhunen *et al.*, 2016). However, the differences between environments are often not only related to a single environmental effect, but combinations of different factors such as hygienic conditions, types of specific diseases, level of pathogen burden, type of feed, feeding regimes, and management of farms. In addition, records on environmental effects such as pathogen burden are not often available. Therefore, it is difficult to find an environmental parameter that explains GxE. To combine effects of climate, feed and management, the estimates of herd-year-season have been used as the environmental parameter for reaction norm analysis of total born in pigs using genomic approach (Silva *et al.*, 2014). Other studies (Herrero-Medrano *et al.*, 2015; Mathur *et al.*, 2014) uses the concept of challenge load and herd-year-week estimates of challenge load as measure of environmental parameters for reaction norm analysis of GxE. Herrero-Medrano *et al.* (2015), Mathur *et al.* (2014) and (Silva *et al.*, 2014) carried out reaction norm analysis in two steps: (i) estimate the environmental parameter from a non-reaction norm model and (ii)

use the estimated environmental parameter as covariates for the reaction norm model. However, this two-step approach may be suboptimal because the reaction norm model treats the environmental parameter as known covariates. The non-reaction norm model may not be the appropriate model, thus the estimated environmental parameter from the non-reaction norm model can be inappropriate values for their use in the reaction norm model. Since a breeding value is defined as a function of the environmental parameter in the reaction norm model, inappropriate inferences about the environmental parameter may have negative consequences on accuracy of EBVs from the model. Su *et al.* (2006) proposed a Bayesian analysis of the reaction norm model that treats the environmental parameter as unknown covariates. This approach, which is done in one step, infers the environmental parameter simultaneously with other parameters of the reaction norm model. Knap and Su (2008) showed considerable advantage of the one-step Bayesian approach of the reaction norm model in pigs. However, the approach by Su *et al.* (2006) is very computationally demanding (Knap and Su, 2008). The use of reaction norm model for selection of robust animals was shown in studies for pigs (Herrero-Medrano *et al.*, 2015; Silva *et al.*, 2014; Zumbach *et al.*, 2008) and cattle (Santana *et al.*, 2016), but to date, no studies on the use of records from multiple environments for selection of robust chicken. This may be due to the difficulty in tracing pedigree for chicken in the commercial conditions.

GxE interactions due to micro-environmental differences

Environments that differ by hygienic conditions, diets, production systems or farms are known as macro-environments. GxE due to differences between macro-environments (environments of purebred versus crossbred animals, and breeding versus production environments) has been discussed. However, within a single macro-environment, micro-environmental differences may also lead to GxE interactions. Each animal may have a different environment e.g. some animals have diseases, but others don't. Different genotypes react differently to the random perturbations that leads to genetic differences in residual variance or genetic heterogeneity of residual variance, i.e. the genotypes may control the residual to some extent.

Genetic heterogeneity of residual variance can be modelled in an additive model, an exponential model or a reaction norm model reviewed in Mulder (2007). An example of the exponential model (SanCristobal-Gaudy *et al.*, 1998) is:

$$y_i = \mu + a_{m,i} + exp\left(\frac{\ln(\sigma_e^2)+a_{v,i}}{2}\right)\chi_i \qquad (7.3)$$

where y_i is phenotypic record of animal *i*; μ is the mean trait value; σ_e^2 is the mean micro-environmental variance; $a_{m,i}$ is the breeding value of animal *i* for the mean: $a_m \sim N(0, \mathbf{A}\sigma_{a_m}^2)$; $a_{v,i}$ is the breeding value of

animal *i* for the micro-environmental variance: a_v ~ N(0, $\mathbf{A}\sigma^2_{a_v}$); **A** is a relationship matrix; $\sigma^2_{a_m}$ and $\sigma^2_{a_v}$ are variances of additive genetic effects of $a_{m,i}$ and $a_{v,i}$; and χ_i is a standard normal deviate for the micro-environmental effect, and it is independent and identically distributed as N[0; 1].

The genetic heterogeneity of residual variance can be estimated using two-stage method (Bolet *et al.*, 2007; Mulder *et al.*, 2009). In stage 1, variance components including additive genetic variance and residual environmental variance are estimated using phenotypic records of individuals. In stage 2, genetic variance in the residual variance is estimated using predicted environmental residual effect from stage 1 as phenotypic records. For example, model for stage 1 is a regular animal model:

$y_i = \mu + a_{m,i} + e_i$ (7.4)

where e_i is residual environmental term of animal *i* with variances of σ^2_e.

Model for stage 2 is similar to a regular animal model, but the model uses log-transformed squared residuals, ln(e_i^2), as phenotypic records. The transformation is to reduce the dependency of e_i^2 on its variance and non-normality of the distribution of e_i^2 (Mulder *et al.*, 2009). The model for stage 2 is:

$\ln(e_i^2) = \mu + a_{v,i} + e_{e_i}$ (7.5)

where $a_{v,i}$ is the breeding value of animal *i* for ln(e_i^2); and e_{e_i} is the residual effect.

Genetic correlation between the additive genetic effects for the mean and the residual variance can be estimated using bivariate model:

$$\begin{bmatrix} \boldsymbol{y} \\ \ln(\boldsymbol{e^2}) \end{bmatrix} = \mathbf{1}\mu + \begin{bmatrix} \boldsymbol{Z_y} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}_{\ln(e^2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{a_m} \\ \boldsymbol{a_v} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e_y} \\ \boldsymbol{e_e} \end{bmatrix} \qquad (7.6)$$

where **y**, **ln(e^2)**, $\mathbf{a_m}$, $\mathbf{a_v}$, $\mathbf{e_y}$ and $\mathbf{e_e}$ are vectors of y_i, ln(e_i^2), $a_{m,i}$, $a_{v,i}$, e_i and e_{e_i}, respectively; $\mathbf{Z_y}$ and $\boldsymbol{Z}_{\ln(e^2)}$ are incidence matrices for the additive genetic effects of $a_{m,i}$ and $a_{v,i}$, respectively; $\begin{bmatrix} \boldsymbol{a_m} \\ \boldsymbol{a_v} \end{bmatrix} \sim MVN(\mathbf{0}, \boldsymbol{A} \otimes \begin{bmatrix} \sigma^2_{a_m} & \sigma_{a_m,a_v} \\ \sigma_{a_m,a_v} & \sigma^2_{a_v} \end{bmatrix})$, where $\sigma^2_{a_m}$ and $\sigma^2_{a_v}$ are variance of $a_{m,i}$, variance of $a_{v,i}$ and covariance of $a_{m,i}$ and $a_{v,i}$.

Covariance between e_i and e_{e_i} is assumed uncorrelated because *cov(e,e^2)* =0 when *e* is normally distributed (Mulder *et al.*, 2009).

Genetic heterogeneity of residual variance has been estimated for BW of broiler chicken (Mulder *et al.*, 2009; Wolc *et al.*, 2009), milk yield in dairy cattle (Rönnegård *et al.*, 2013), production traits in beef cattle (Neves *et al.*, 2012). Generally, these studies found that estimates of heritability of residual variation were low at 0.023-0.038 (Wolc *et al.*, 2009), 0.029-0.047 (Mulder *et al.*, 2009) and 0.00-0.05 (Neves *et al.*, 2012). The genetic correlations between the additive genetic effects for the mean and the residual variance were ranging from -0.45 to -0.41 (Mulder *et al.*, 2009). Although heritability of residual variation was low, the

genetic coefficients of variation were relatively high ranging at 0.35-0.57 (Mulder *et al.*, 2009) and 0.25-0.40 (Wolc *et al.*, 2009). This indicates that accounting for genetic heterogeneity of residual variance in selection can improve uniformity of the traits of interest.

In summary, to improve poultry breeding programs affected by GxE, the use of records from purebred chicken in C is more reasonable than the use of records from crossbred chicken in C, because the use of records from purebred in C has higher accuracy of selection, implementation of breeding schemes is less complicated, and chicken may have relatively low GxG interactions due to crossbreeding. Research on selection for multiple production environments has not been well-investigated for poultry breeding programs.

7.2. Development of statistical models for genetic evaluations

A true model that describes the pattern of the data for a trait of interests is usually unknown because of missing information, incorrect assumptions and computational issues. Since the true model is unknown, there are controversies on how to develop a statistical model for genetic evaluations of traits. For example, some studies use statistic measures of model goodness-of-fit such as the deviance information criterion (DIC), likelihood-ratio test (LRT) and mean square error (MSE) (Grosso *et al.*, 2010; Kheirabadi and Rashidi, 2019; Posht-e Masari *et al.*, 2019). A common method for development of the model, particularly used as norm in the development of genomic prediction models, is cross-validation for assessing predictive ability of EBV prediction from statistical models (Christensen *et al.*, 2012; Gorjanc *et al.*, 2015; Legarra and Reverter, 2018). Different cross-validation approaches to measure accuracy of EBV prediction have been used such as population accuracy based on correlation between EBVs and corrected phenotypes (Christensen *et al.*, 2012), and indicators of population accuracy based on correlation between EBVs estimated from reduced and full datasets (Legarra and Reverter, 2018). In addition, appropriate approaches to develop statistical models depend on the traits. For example, predictive ability based on correlation between EBVs and corrected phenotypes may not be appropriate for the traits that are affected by maternal effects (Chapter 4). Also, corrected phenotypes of the traits measured in C environment may not be available for validation animals that are raised in B environment. This section will list some methods and approaches for the development and comparison of statistical models.

Model goodness-of-fit

Goodness-of-fit of a model indicates how well the model describes a set of data, which is an important component of model comparison. There are several common criteria to assess the goodness-of-fit of a model such as Akaike information criterion (AIC), Bayesian information criterion (BIC) and deviance

information criterion (DIC). Criteria AIC and BIC are based on the likelihood function for frequentist-oriented approach. The AIC is an unbiased estimator of the Kullback-Leibler divergence of the evaluated model from the true model (Akaike, 1973). The BIC is closely related to AIC, but BIC was developed by Schwarz (1978) which used a Bayesian argument for adopting it. The criteria AIC and BIC used the penalized-likelihood approach that includes measure of the fit of a model and an additional term that penalizes the complexity of the model:

$AIC = -2\log(f(y|\hat{\theta})) + 2k$ (7.7)

$BIC = -2\log(f(y|\hat{\theta})) + 2k \log(n)$ (7.8)

where y is the data; $\hat{\theta}$ is the vector of the model parameter estimates at the maximum likelihood of $f(y|\hat{\theta})$; k is the number of estimated parameters in the model; n is the number of data observations.

The BIC penalizes the complexity of the model further than the AIC. The model with smaller criterion of AIC or BIC is the better model. The use of AIC and BIC values leads to similar conclusions on the choice of statistical models for poultry traits (Grosso *et al.*, 2010; Jasouri *et al.*, 2017; Maniatis *et al.*, 2013).

The maximized value of the likelihood function $f(y|\hat{\theta})$, or called L, is also used to compare hierarchically nested models in a log-likelihood ratio test. The test is based on the Chi-Square test of the ratio of the likelihood of a relatively more complex model compared to a simpler model: $LR = -2\log L_2 + 2\log L_1$, where $\log L_1$ is the log-likelihood of the complex model and $\log L_2$ is the log-likelihood of the simple model. The LR value is compared with the critical value obtained from the Chi-Square distribution at usually $P=0.05$ and degrees of freedom = $n_{k1} - n_{k2}$, where n_{k1} and n_{k2} are number of estimated parameters of the complex model and simpler model, respectively. The complex model and simple model are significantly different ($P=0.05$) if LR is greater than the critical value from a Chi-Square distribution. In addition, the likelihood ratio test is used to test the significance of random factors. To test the significance of random factors, the Wald test can also be used. The Wald test requires estimates from only one model. For example, in the test for a single parameter $\hat{\theta}$ with H_o: $\hat{\theta} = \theta_0$ and H_1: $\hat{\theta} \neq \theta_0$, the Wald statistic is $\frac{(\hat{\theta}-\theta_0)^2}{var(\hat{\theta})}$, where θ_0 is a constant. The Wald statistic is compared with a chi-shared distribution at degree of freedom of 1.

Criterion DIC (Spiegelhalter *et al.*, 2002) is a Bayesian analog of AIC defined as:

$DIC = -2\log(f(y|\bar{\theta})) + 2p_D$ (7.9)

where $f(y|\bar{\theta})$ is the likelihood evaluated at the posterior mean $\bar{\theta}$ (i.e. mean of the posterior distribution); and p_D is a model complexity measure, known as the effective number of parameters (Myung and Pitt, 2018).

Formula (7.7) and (7.9) are quite similar except that $\hat{\theta}$ is replaced by $\bar{\theta}$, and *k* is substituted by p_D. The criterion DIC may be sensitive to prior distributions of model parameters.

Other measures of the goodness-of-fit of statistical models are mean square error (MSE) and correlation between observed and predicted values (Posht-e Masari *et al.*, 2019). These measures can be used for both frequentist and Bayesian approaches. The measures are: $MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$ and $r_{(\mathbf{y},\hat{\mathbf{y}})} = \frac{cov(\mathbf{y},\hat{\mathbf{y}})}{\sqrt{\sigma_y^2 \sigma_{\hat{y}}^2}}$,

where *MSE* is mean square error; $r_{(\mathbf{y},\hat{\mathbf{y}})}$ is the correlation between observed and predicted phenotypic values; y_i is the observed value of phenotypic record *i*; $\hat{y}_i$ is the value of phenotypic record *i* predicted from the statistical model; *n* is number of records; $cov(\mathbf{y},\hat{\mathbf{y}})$ is the covariance between observed and predicted values; σ_y^2 and $\sigma_{\hat{y}}^2$ are variances of observed and predicted values, respectively. Models with *MSE* closer to 0 and $r_{(\mathbf{y},\hat{\mathbf{y}})}$ closer to 1 are the better models. The use of DIC, MSE and $r_{(\mathbf{y},\hat{\mathbf{y}})}$ led to similar conclusions on the goodness-of-fit between comparable statistical models in a study by Posht-e Masari *et al.* (2019).

Accuracy of estimated breeding values

Identification of statistical models that fit data well is important. However, for breeders, accuracy of prediction of breeding values is more important in the development of statistical models. Genetic gain of a breeding program depends on accuracy of EBVs: $R = i\rho\sigma_A$, where R is selection response; ρ is accuracy of EBVs or the correlation between true breeding values and EBVs for selection candidates; σ_A is the additive genetic standard deviation of the trait of interest. A common method to calculate the accuracy is from the prediction error variance (PEV) and the additive genetic variance:

$$\rho_{PEV} = \sqrt{1 - \frac{PEV}{\sigma_A^2}} \qquad (7.10)$$

where PEV is the prediction error variance derived from the diagonal elements of the inverse left-hand side matrix of MME; and σ_A^2 is the additive genetic variance of the trait.

However, in formula (7.10), $1 - \frac{PEV}{\sigma_A^2}$ could be negative if there was inbreeding. Therefore, the additive genetic variance should be multiplied with the diagonal element of the additive numerator relationship matrix (**A**). The accuracy of EBVs derived from PEV is known as individual accuracy (Bijma, 2012; Legarra and Reverter, 2018), theoretical accuracy (Mrode and Thompson, 2005; Putz *et al.*, 2018) or model-based accuracy. The individual accuracy is defined as the correlation ($r(a,\hat{a})$) between the true breeding values (**a**) and estimated breeding values (**â**) for one individual across conceptual repeated sampling (Legarra and Reverter, 2018; Van Vleck, 1993). The individual accuracy reflects "the credibility of an individual EBV" or "standard error of prediction of an individual EBV" (Bijma, 2012), and the individual accuracy "relates to

the risk that this EBV will change over time when more information becomes available" (Bijma, 2012). However, the individual accuracy calculated from PEV ignores selection while selection reduces the accuracy (Bijma, 2012). In addition, the individual accuracy might be meaningless to selection. For example, when only parent average is known, individual accuracy of full-sibs is at most 0.71. However, the predicted differences between full sibs have zero accuracy with the pedigree-based BLUP model, or population accuracy among full-sibs is zero because full-sibs have the same parent average. Therefore, individual accuracy derived from PEV should be used for individual decisions, supplementary information for customers and marketing purposes, but the individual accuracy might not be appropriate for selection, genetic gain, and the development of statistical models (Bijma, 2012).

The use of population accuracy of EBVs may be more appropriate for a comparison of competing statistical models (Legarra and Reverter, 2018). Population accuracy is defined as the correlation ($\rho(\mathbf{a},\hat{\mathbf{a}})$) between the true breeding values and estimated breeding values across series of individuals in a population or across the candidates for selection: $\rho(\mathbf{a},\hat{\mathbf{a}}) = \frac{cov(\mathbf{a},\hat{\mathbf{a}})}{\sqrt{var(\mathbf{a})var(\hat{\mathbf{a}})}}$, where $cov(\mathbf{a},\hat{\mathbf{a}})$ is the covariance between the true breeding values ($\mathbf{a}$) and estimated breeding values $\hat{\mathbf{a}}$ across individuals in a population; $var(\mathbf{a})$ and $var(\hat{\mathbf{a}})$ are variances of $\mathbf{a}$ and $\hat{\mathbf{a}}$, respectively. The population accuracy of EBVs is the property of a population, not of an individual (Bijma, 2012; Legarra and Reverter, 2018). In the absence of selection, individual accuracy can be equal to population accuracy (Bijma, 2012). To estimate the population accuracy of EBVs, the common method is based on cross-validation, which is a norm in genomic evaluation studies (Christensen *et al.*, 2012; Legarra and Reverter, 2018). The term, accuracy of EBVs, used in this thesis generally refers to or being indicators of population accuracy of EBVs unless a specific form of accuracy is stated.

Cross-validation

Cross-validation is a model validation test to assess the ability of the model to predict future events, typically based on predicted breeding values. In a cross-validation analysis, the full dataset is split into a training set (or reduced dataset) and a validation set (or testing dataset). Breeding values of animals in the validation dataset are predicted from the training set. The predicted breeding values (EBVs) are compared with target predictands (observed values) that are derived from the prediction based on the full dataset. Methodologies of cross-validation vary across studies for approaches of dataset splitting, compared predictands and metrics for the prediction of EBVs. The difference in the methodologies of cross-validation can be due to characteristics of traits/effects, breeding schemes, statistical models and target individuals (selection candidates).

Splitting of the full dataset can be based on a cut-off date (Christensen *et al.*, 2012), random folding (Su *et al.*, 2010), split folds across families (Legarra *et al.*, 2008) and split folds within families (Chu et al. 2019), which can lead to very different results of accuracy of EBVs (Daetwyler *et al.*, 2013; Legarra and Reverter, 2017). Forward cross-validation uses a cut-off date to split the full dataset into training and validation datasets (Mäntysaari *et al.*, 2010). The cut-off date is not in any form of randomization or related to measure of uncertainty in data splitting. The idea of forward cross-validation is to use the "current" data to forecast the "future" outcome. Random folding *k*-fold cross-validation splits randomly the full data into *k* distinct sets and predict EBVs of animals in a set from the remaining *k-1* sets. Random folding cross-validation is relatively simple for implementation (Legarra and Reverter, 2017), but random folding cross-validation is not realistic in an animal breeding setting in which we'd like to forecast future outcome of breeding. Splitting of the full dataset into training and validation sets across families and within families has been used in Legarra *et al.* (2008). Chu et al. 2019 split the full dataset based on combined approaches of forward cross-validation and split folds within families. The split folds within families in Chu et al. 2019 aimed to amplify the overestimation of direct additive genetic effects when maternal effects were not accounted for in the model. Data splitting methods in cross-validation procedures can be varied across studies, but the principle of generating training and validation sets should be corresponding to testing hypothesis and should mimic the relationship of the candidates of selection to the training population (Daetwyler *et al.*, 2013). The relationship can affect accuracy of EBVs. The accuracy would be higher when the relationship between animals of the training and validation sets was higher (Daetwyler *et al.*, 2013).

The true breeding values are usually unknown, so is the true accuracy of EBVs, $r(\mathbf{a},\hat{\mathbf{a}})$. To estimate the accuracy, EBVs estimated from the training dataset are compared with predictands such as pre-corrected (or observed) phenotypes (Christensen *et al.*, 2012; Legarra *et al.*, 2008), yield deviations, daughter yield deviations (Mäntysaari *et al.*, 2010), de-regressed proofs (Legarra *et al.*, 2008) and EBVs estimated from the full dataset (Legarra and Reverter, 2018). Yield deviations, daughter yield deviation and de-regressed proofs are pseudo-phenotypes commonly used in cattle breeding programs, but these predictands have not been used in cross-validation studies for chicken breeding programs, in which prediction of EBVs is primarily based on own performance and performance from siblings. Corrected phenotypes are adjusted for fixed effects (Legarra *et al.*, 2008) or adjusted for fixed effects and non-genetic random effects (Christensen *et al.*, 2012). The corrected phenotypes are derived from a model that uses the full dataset. This means corrected phenotypes are estimated values, and they can be biased due to the use of an incorrect model or wrongly assumed relationship matrix (Legarra and Reverter, 2017). Cross-validation approach that is based on the predictands of corrected phenotypes is sensitive to incorrect estimates of

heritability (Legarra and Reverter, 2018). In addition, corrected phenotypes may not exist in some cases such as breeding programs affected by GxE. In these cases, predictands of EBVs that are estimated from the full dataset can be compared with EBVs that are estimated from the training dataset (Legarra and Reverter, 2018). This cross-validation procedure is relatively simple tool for evaluation of competing statistical models. Legarra and Reverter (2018) found that the use of predictands of EBVs estimated from the full dataset led to similar conclusion on the comparison of models as the use of corrected phenotypes. However, the two models compared in Legarra and Reverter (2018) were pedigree-based and genomic-based models which have substantial differences in accuracy of EBVs. Therefore, it is uncertain that the predictands of EBVs estimated from the full dataset can be used for selection of statistical models when the differences in accuracy of EBVs between competing models are relatively small. In addition, EBVs from the full dataset are estimates which can be biased due to the use of incorrect models.

Common metrics for the prediction of EBVs from cross-validation procedure are accuracy of EBVs, bias, dispersion and mean square error. Since true breeding values are unknown, estimation of accuracy of EBVs is usually based on predictands. Metrics of accuracy of EBVs vary across studies (Christensen *et al.*, 2012; Legarra and Reverter, 2018; Legarra *et al.*, 2008; Putz *et al.*, 2018):

$$\rho_{\widehat{y_c}} = cor(\widehat{\mathbf{a}}_\mathbf{r}, \widehat{\mathbf{y}}_\mathbf{c}) \qquad (7.11)$$

$$\rho_{\widehat{y_c},h} = \frac{cor(\widehat{\mathbf{a}}_\mathbf{r}, \widehat{\mathbf{y}}_\mathbf{c})}{\sqrt{h^2}} \qquad (7.12)$$

$$\rho_{\widehat{a_f}} = cor(\widehat{\mathbf{a}}_\mathbf{r}, \widehat{\mathbf{a}}_\mathbf{f}) \qquad (7.13)$$

where $\widehat{\mathbf{a}}_\mathbf{r}$ is the vector of EBV estimated from the training or reduced dataset; $\widehat{\mathbf{a}}_\mathbf{f}$ and $\widehat{\mathbf{y}}_\mathbf{c}$ are vectors of EBV and corrected phenotype, respectively estimated from the full dataset; *cor* is the correlation; h^2 is heritability estimate; $\rho_{\widehat{y_c}}$, $\rho_{\widehat{y_c},h}$ and $\rho_{\widehat{a_f}}$ are metrics of accuracies of EBVs used in different studies. Metric $\rho_{\widehat{a_f}}$ has expectation of $\frac{acc_r}{acc_f}$ (Legarra and Reverter, 2018), where acc_r is the population accuracy of EBVs defined as the correlation between the true breeding values and EBVs estimated from the training dataset; acc_f is the population accuracy of EBVs defined as the correlation between the true breeding values and EBVs estimated from the full dataset.

A study (Putz *et al.*, 2018) that simulated a population without selection compared accuracies of EBVs estimated from pedigree based BLUP and single step GBLUP models. Using different approaches to calculate accuracy of EBVs, Putz *et al.* (2018) found that the values of $\rho_{\widehat{y_c},h}$ and individual accuracy were close to the correlation $\rho(a, \hat{a})$ between estimated breeding values and true breeding values (Table 7.1).

Table 7.1: Mean (standard deviation over 10 replicates) of accuracies of breeding values estimated from pedigree-based BLUP and single step BLUP (ssGBLUP) using the training dataset (Table adapted from Putz *et al.* (2018))

Notation from Putz *et al.* (2018)	Notation in this manuscript	EBVs from pedigree-based BLUP	GEBVs from ssGBLUP
TBV[a]	$\rho(\mathbf{a}, \hat{\mathbf{a}})$	0.34 (0.06)	0.44 (0.06)
PEV[b]	ρ_{PEV}	0.30 (0.02)	0.46 (0.01)
$GEBV_{full}$[c]	$\rho_{\widehat{a_f}}$	0.56 (0.06)	0.82 (0.03)
Y_c[d]	$\rho_{\widehat{y_c}}$	0.11 (0.05)	0.14 (0.03)
$Y_{c,h}$[e]	$\rho_{\widehat{y_c},h}$	0.36 (0.15)	0.45 (0.11)
$Y_{c,s}$[f]	$\rho_{\widehat{y_c}}$	0.14 (0.22)	0.17 (0.22)

Notes: [a] Correlation between estimated breeding values and true breeding values; [b] Individual accuracy computed from prediction error variances; [c] Correlation between estimated breeding values and genomic breeding values estimated from the full dataset; [d] Correlation between estimated breeding values and corrected phenotypes for dams; [e] Y_c divided by the square root of heritability estimates; [f] Correlation between estimated breeding values and average corrected phenotypes of daughters for sires with at least five daughters.

Metrics of $\rho_{\widehat{y_c}}$ and $\rho_{\widehat{a_f}}$ did show that prediction from ssGBLUP was better than pedigree-based BLUP. However, the correlation, $\rho_{\widehat{y_c}}$, between estimated breeding values and corrected phenotypes was significantly different from the true accuracy of EBVs. Similarly, the correlation, $\rho_{\widehat{a_f}}$, between estimated breeding values estimated from the training dataset and genomic breeding values estimated from the full dataset was significantly different from the true accuracy of EBVs. This is because $\rho_{\widehat{y_c}}$ and $\rho_{\widehat{a_f}}$ are not accuracy of EBVs, but $\rho_{\widehat{y_c}}$ and $\rho_{\widehat{a_f}}$ are ratios of accuracies that describes the change in accuracy of EBVs from the use of the training dataset to the full dataset. Computation of accuracy of EBVs from $\rho_{\widehat{a_f}}$ requires prediction error variances and covariances, and genetic variance at equilibrium in a population under selection (Legarra and Reverter, 2018).

Other metrics for model comparisons are bias, dispersion and mean square error. Bias is defined as the mean of the difference between true breeding value and EBV: $d = \bar{a} - \bar{\hat{a}}$, where *d* is bias of prediction; $\bar{a}$ is the mean of true breeding values, and $\bar{\hat{a}}$ is the mean of estimated breeding values. Dispersion is defined as the regression slope (*b*) of **a** on $\hat{\mathbf{a}}$. Dispersion is defined as bias in some studies (Christensen *et al.*, 2012), but *b* actually describes the inflation or deflation of EBV prediction (Legarra and Reverter, 2018).

Mean square error is defined as the mean of the square of the difference between true breeding value and EBV: *MSE* = $\overline{(\mathbf{a}-\hat{\mathbf{a}})^{\mathbf{2}}}$. For estimation of bias, dispersion and mean square error from cross-validation procedures, predictants such as corrected phenotypes and EBVs estimated from the full dataset are used in replacement for the unknown true breeding values.

Here are three examples of cross-validation procedures used in three different studies:

Christensen *et al.* (2012) used forward cross-validation procedures and corrected phenotypes as predictands. The full dataset was split into a training dataset and a validation dataset based on a cut-off date, and the validation dataset was ensure to have animals with both parents in the training dataset. The cross-validation in Christensen *et al.* (2012) aimed to compare accuracy of EBVs between pedigree-based BLUP and ssGBLUP models. The training dataset was used for estimation of model parameters and predicting EBVs of animals in the validation dataset for the different models. The metric of accuracy of EBVs was reflected by the correlation ($\rho_{\widehat{y_c}}$) between EBVs and corrected phenotypes of the same animals in the validation dataset: $\rho_{\widehat{y_c}} = cor(\widehat{\mathbf{a}_{\mathbf{r,1}}}, \widehat{\mathbf{y}_{\mathbf{c,1}}})$, where $\widehat{\mathbf{a}_{\mathbf{r,1}}}$ is EBV of animal *i* estimated from the training dataset using either pedigree-based BLUP or ssGBLUP model; $\widehat{\mathbf{y}_{\mathbf{c,1}}}$ is corrected phenotype of animal *i* estimated from the full dataset using pedigree-based BLUP model. The corrected phenotypes were adjusted for fixed effects and non-genetic random effects. The dispersion of prediction was assessed by the regression slope of $\widehat{\mathbf{y}_{\mathbf{c,1}}}$ on $\widehat{\mathbf{a}_{\mathbf{r,1}}}$.

Chapter 4 in this thesis, used cross-validation procedures that were based on half-sib correlation. The full dataset was split into training datasets and validation datasets based on cut-off date and full-sib families. Animals in the validation datasets was ensured to have full-sibs in the training datasets. Randomization of data splitting was accounted for as there were many possibilities of allocating full-sibs into training datasets and validation datasets. The full dataset was used to estimate model parameters and corrected phenotypes of animals for different models. Competing models in the study of Chapter 4 were pedigree-based statistical models that included different fixed effects and random effects. The metric of accuracy of EBVs $\rho_{\widehat{y_c},h}$ was computed as: $\rho_{\widehat{y_c},h} = \frac{cor(\widehat{\mathbf{a}_{\mathbf{r,1}}}, \widehat{\mathbf{y}_{\mathbf{c,j}}})}{\sqrt{h^2 r_{i,j}}}$, where $\widehat{\mathbf{a}_{\mathbf{r,1}}}$ is EBV of animal *i* estimated from the training dataset; $\widehat{\mathbf{y}_{\mathbf{c,j}}}$ is the corrected phenotype of animal *j* estimated from the full dataset; animals *i* and *j* in the validation datasets are half-sibs that are offspring from the same sire, but different dams; $r_{i,j}$ is the additive genetic relationship between half-sibs that is assumed to be equal to 0.25; h^2 is heritability estimate from the full dataset. The corrected phenotypes were adjusted for fixed effects. The cross-validation procedures used in Chapter 4 were designed for traits that might be affected by maternal effects. When the maternal effects are present, but not accounted for appropriately in the model, the

prediction of EBVs of animals in the validation dataset may be overestimated due to the maternal effects shared among full-sibs. To avoid this overestimation, $\rho_{\widehat{y_c},h}$ was based on the correlation between different animals that are half-sibs from different dams, and due to the mating structure such half-sibs will not share potential bias from maternal effects.

Legarra and Reverter (2018) compared accuracy of EBV prediction for a pedigree-based BLUP model and a GBLUP model using different statistics that were derived from cross-validation procedures. The full dataset was randomly split with 50% and 50% of records into training datasets and validation datasets, respectively. The full dataset in Legarra and Reverter (2018) was from a population of animals that are not descendants from each other. The full dataset and randomly generated training datasets were used to estimate model parameters and EBVs of animals for the different models. The corrected phenotypes were estimated from the full dataset for the different models. Different statistics for accuracy, bias and dispersion of EBVs was computed based on predictands of corrected phenotypes or EBVs estimated from the full dataset. Metrics of accuracy of EBVs were the correlation ($\rho_{\widehat{a_f}}$) between EBVs estimated from the full dataset and the validation dataset and the correlation ($\rho_{\widehat{y_c}}$) between EBVs estimated from the validation dataset and corrected phenotypes. Those correlations was between estimates of the same animals in the validation datasets. Interestingly, Legarra and Reverter (2018) concluded that $\rho_{\widehat{a_f}}$ performed better than $\rho_{\widehat{y_c}}$ as the metric of accuracy of EBVs $\rho_{\widehat{a_f}}$ was more closely related to the changes in prediction of EBVs. The metric $\rho_{\widehat{a_f}}$ was used in Chapter 2 to assess accuracy of EBVs for a pedigree-based BLUP model and a GBLUP model. The cross-validation method in Legarra and Reverter (2018) is relatively simple to implement and particularly useful in case of GxE breeding programs in which corrected phenotypes are unavailable for validation animals or selection candidates.

In summary, there are number of tools and approaches for evaluating statistical models. However, there are lack of comparisons between model evaluation methods, and it is little known which evaluation method is best for traits of interests in poultry breeding programs.

7.3. Group records

It is shown in Chapter 3 and 5 that when GxE interactions between breeding and production environments are strong, information from the production environment is a key factor to improve genetic gain. However, for traits like egg production and feed intake, continuous recording of individual phenotypes is difficult in C environments. For example, records of egg production are only available in groups such as hen housed egg production in village production systems (Wondmeneh *et al.*, 2016) and group records of

cages in commercial laying hens. Intake of feed, which typically represents 60-70% of production costs, are important traits, but difficult or expensive to measure. Since individual recording of feed intake is expensive, only a selective proportion of selection candidates is tested for feed efficiency traits (Mebratie, 2019). For the genetic progress of feed efficiency, the lack of phenotypic data on the traits can be a barrier. Group records can be an attractive alternative to individual records. In addition, chicken are typically caged individually in the recording systems of breeding programs for selection of feed efficiency traits whereas in the commercial conditions, birds are housed in groups. The difference in housing conditions between the breeding and production environments may lead to GxE interactions. For example, Zerehdaran *et al.* (2005) showed that the genetic correlation of broiler BW between birds housed in groups and individual cages was 0.80. Therefore, compared to individual records, the advantages of group records are that group records are easier and cheaper to collect (in some cases, the data is only available in records of groups), and group records may reflect better the commercial conditions of the production environment.

For estimation of genetic parameters and prediction of EBVs from group records, there are three model approaches. The model approach proposed in Olson *et al.* (2006), Biscarini *et al.* (2010) and Su *et al.* (2018), let's call M_{gr_exact}, uses the sum of group records and assumes only one residual for each phenotypic group record. Olson *et al.* (2006) used model M_{gr_exact} (known the "exact method") to predict EBVs in case of equal group sizes (number of animals per group), but variance components were not estimated. Model M_{gr_exact} for group records used in Su *et al.* (2018) could estimate variance components of additive genetic effects, group effects, maternal permanent environmental effects and residuals that were consistent with the variances estimated from individual records. Model M_{gr_exact} in Biscarini *et al.* (2010) and Chapter 6 in this thesis applied for equal group sizes also resulted in variance components of additive genetic effects and residuals that were consistent with the variances estimated from individual records. The model approach for group records, termed $M_{gr\_average}$, uses the average of group records as a replacement for the phenotypic records of individuals. The $M_{gr\_average}$ approach, which is known as the approximate method in Olson *et al.* (2006) and Biscarini *et al.* (2010), has a relatively simpler setting of MME and easier implementation compared to M_{gr_exact} approach. When variance components were known and given, accuracy of EBVs from $M_{gr\_average}$ approach was only slightly smaller than M_{gr_exact} approach (Olson *et al.*, 2006). The Pearson correlation and rank correlation between EBVs estimated from $M_{gr\_average}$ and M_{gr_exact} approaches were high (0.86-0.91) (Biscarini *et al.*, 2010). However, variance components estimated from $M_{gr\_average}$ approach were very different from the variances estimated from M_{gr_exact} approach (Biscarini *et al.*, 2010) and different from the variances estimated from individual records (unpublished results from

my own simulations for Chapter 6 of this thesis). Variance components estimated from $M_{gr\_average}$ approach depended on structure of groups (unpublished results from my own simulations). The third approach is, termed M_{gr_sire}, used in Nurgiartiningsih *et al.* (2004) that is based on the sire model. The M_{gr_sire} approach can only be applied to group records in which group members are full-sibs and half-sibs. The M_{gr_sire} approach incorporates heterogeneity of error variance and correlated residual effects into the model as the residual of this sire model consists of the group effect, dam effect and environmental terms. Nurgiartiningsih *et al.* (2004) showed that using M_{gr_sire} approach, the additive genetic variance estimated from group records was, generally, consistent with the additive genetic variance estimated from individual records. The rank correlation between breeding values of sires predicted from group records and individual records was 0.79-0.99 (Nurgiartiningsih *et al.*, 2004). However, compared to the prediction from individual records, the residual variances were largely overestimated, thus heritability estimated from group records was significantly lower than heritability estimated from individual records (Nurgiartiningsih *et al.*, 2004). The overestimation of the residual variances might be mainly due to confounding of the group effect into the residual.

Models of $M_{gr\_average}$ and M_{gr_exact} for group records, which were used in Chapter 5 and 6, respectively, will be described and compared with the model for individual records. These models include fixed effects and random effects of additive genetics, maternal permanent environment and group):

Model for individual records: $\mathbf{y} = \mathbf{Xb} + \mathbf{Z_a a} + \mathbf{Z_{gr} c_{gr}} + \mathbf{Z_m c_m} + \mathbf{e}$ (M_{ind})

Model $M_{gr\_average}$ for group records: $\mathbf{y^t} = \mathbf{Xb} + \mathbf{Z_a a} + \mathbf{Z_{gr} c_{gr}} + \mathbf{Z_m c_m} + \mathbf{e}$ ($M_{gr\_average}$)

Model M_{gr_exact} for group records: $\mathbf{y^*} = \mathbf{X^* b} + \mathbf{Z_a^* a} + \mathbf{Z_{gr}^* c_{gr}} + \mathbf{Z_m^* c_m} + \mathbf{e^*}$ (M_{gr_exact})

where $\mathbf{y}$ is a vector of individual phenotypic records; $\mathbf{y^t}$ is a vector of individual records, but the individual records are group means; $\mathbf{y^*}$ is a vector of group records that are sum records of individual in groups. Vectors $\mathbf{y}$ and $\mathbf{y^t}$ have the same number of elements that are equal to number of individual phenotypic records (n_{ind}). Vectors $\mathbf{y^*}$ has number of elements equal to number of groups (n_{gr}). Vector $\mathbf{b}$ is a vector of fixed effects; $\mathbf{a}$ is a vector of direct additive genetic effects: $\mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2)$, where $\mathbf{A}$ is a relationship matrix between individuals constructed from pedigree, genomic information or combination of pedigree and genomic information, σ_a^2 is the direct additive genetic variance; $\mathbf{c_{gr}}$ is a vector of random group effects: $\mathbf{c_{gr}} \sim N(0, \mathbf{I_{gr}}\sigma_{gr}^2)$, where $\mathbf{I_{gr}}$ is an identity matrix of group effects, σ_{gr}^2 is the group variance; $\mathbf{c_m}$ is a vector of permanent environment maternal effects: $\mathbf{c_m} \sim N(0, \mathbf{I_m}\sigma_m^2)$, where $\mathbf{I_m}$ is an identity matrix of the dam, σ_m^2 is the permanent environment maternal variance; $\mathbf{X}$, $\mathbf{X^*}$, $\mathbf{Z_a}$, $\mathbf{Z_a^*}$, $\mathbf{Z_{gr}}$, $\mathbf{Z_{gr}^*}$, $\mathbf{Z_m}$ and $\mathbf{Z_m^*}$ are incidence matrices. Matrices $\mathbf{X}$ and $\mathbf{X^*}$ have the same number of columns, but $\mathbf{X}$ and $\mathbf{X^*}$ have n_{ind} and n_{gr} number of rows, respectively. Matrices $\mathbf{Z_a}$ and $\mathbf{Z_a^*}$ have the same number of columns equal to number of animals in the

relationship matrix, but $\mathbf{Z_a}$ and $\mathbf{Z_a^*}$ have n_{ind} and n_{gr} number of rows, respectively. Matrices $\mathbf{Z_{gr}}$ and $\mathbf{Z_{gr}^*}$ have n_{gr} number of columns, but $\mathbf{Z_{gr}}$ and $\mathbf{Z_{gr}^*}$ have n_{ind} and n_{gr} number of rows, respectively. Matrices $\mathbf{Z_m}$ and $\mathbf{Z_m^*}$ have the same number of columns equal to number of dams, but $\mathbf{Z_m}$ and $\mathbf{Z_m^*}$ have n_{ind} and n_{gr} number of rows, respectively. Vector **e** is the residuals of models M_{ind} and $M_{gr\_average}$: $\mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$, where **R** is an identity matrix of individuals, σ_e^2 is the residual variance. Vector $\mathbf{e^*}$ is the residuals of model M_{gr_exact}: $\mathbf{e^*} \sim N(0, \mathbf{R^*}\sigma_e^2)$, where $\mathbf{R^*}$ is a diagonal matrix in which diagonal elements are number of animals in groups.

Pedigree-based relationship and genomic-based relationship have been used for the prediction of EBVs. When pedigree-based relationship matrix is used for models $M_{gr\_average}$ and M_{gr_exact} (Biscarini *et al.*, 2010; Olson *et al.*, 2006; Peeters *et al.*, 2013; Su *et al.*, 2018), full-sibs of the same group have identical EBV. On the contrary, when genomic relationship matrix is used for models $M_{gr\_average}$ and M_{gr_exact}, full-sibs of the same group can have different genomic EBVs (Chapter 6). For prediction from group records, the use of genomic information can not only increase accuracy of EBV, but also can reduce inbreeding compared to the use of pedigree information (Chapter 6).

Simulations in Chapter 5 and 6 did not include the permanent environmental maternal effect in the simulation of trait observations. However, the maternal effect can be important for some traits in poultry, particularly for traits that are measured at relatively early age. When the maternal effect is present, but ignored in the prediction model, the direct additive genetic variance is overestimated. Su *et al.* (2018) showed that with the model that ignored the maternal effect, the permanent environmental maternal variance was "transferred" to the direct additive genetic variance, group variance and residual variance. This "transfer" reduced slightly accuracy of EBVs and increased considerably bias of prediction (Su *et al.*, 2018). To be able to estimate the permanent environmental maternal effect from group records, offspring of a dam must belong to different groups, or offspring of multiple dams must belong to the same group. When group size is equal to family size in hierarchical mating and offspring of each dam belong to each group, the permanent environmental maternal effect cannot be estimated from group records.

The variance due to group effects cannot be estimated from group records when group sizes are equal between groups. When the group effect is present, but ignored in the prediction model, the direct additive genetic variance is unaffected, but the residual variance is largely overestimated. Ignoring the group effect reduces slightly accuracy of EBVs, but has no influence on bias of prediction. With the equal group sizes, the group and residual effects cannot be separated because the covariance matrix for the group effect is proportional to the covariance matrix of residuals, (matrices $\mathbf{Z_{gr}^*Z_{gr}^{*\prime}} = n_g\mathbf{R}^*$, where n_g is the group size in model M_{gr_exact}). In this case, the residual variance are: $\sigma_{e^*}^2 = n_g(n_g\sigma_{gr}^2 + \sigma_e^2)$, where $\sigma_{e^*}^2$ is the residual variance in the model that the group effect are ignored; n_g is the group size; σ_{gr}^2 and σ_e^2 are residual and

group variances, respectively. When model M_{gr_exact} without group effect was used in Biscarini *et al.* (2008), residual variances ($\sigma^2_{e^*}$) estimated from group records were much higher than the residual variances (σ^2_e) estimated from individual records. Biscarini *et al.* (2008) assumed the residual variances ($\sigma^2_{e^*}$) estimated from group records must be divided by group sizes ($\sigma^2_e = \frac{\sigma^2_{e^*}}{n_g}$) without realizing that the group effect was incorporated into the residuals. As a consequence, Biscarini *et al.* (2008) actually overestimated σ^2_e for the estimation from group records.

Small variation of group sizes can also lead to convergence issues of the model that includes the group effect for parameter estimation from group records (Su *et al.*, 2018). Therefore, with equal group sizes or small variation of group sizes, Su *et al.* (2018) suggested an alternative model ($M^{c_{gr}}_{gr\_exact}$) that combines the group effect into residuals, and appropriated weights are put on the residuals:

$\mathbf{y}^* = \mathbf{X}^*\mathbf{b} + \mathbf{Z}^*_{\mathbf{a}}\mathbf{a} + \mathbf{Z}^*_{\mathbf{m}}\mathbf{c}_{\mathbf{m}} + \mathbf{e}^*$ ($M^{c_{gr}}_{gr\_exact}$)

Residual variance of $M^{c_{gr}}_{gr\_Su}$ is: $\sigma^2_{e^*} = n_g\left(n_g\sigma^2_{gr} + \sigma^2_e\right) = n_g\left(n_g b + 1\right)$, where $b = {\sigma^2_{gr}}/{\sigma^2_e}$. Thus, $V(\mathbf{e}^*) = \mathbf{R}^*\sigma^2_e$, where $\mathbf{R}^*$ is a diagonal matrix with diagonal elements $R^*_{ii} = n_{gi}\left(n_{gi}b + 1\right)$. Model $M^{c_{gr}}_{gr\_exact}$ is equivalent to model M_{gr_exact}, but $M^{c_{gr}}_{gr\_exact}$ has less computation demand (Su *et al.*, 2018). For model $M^{c_{gr}}_{gr\_exact}$, the ratio $b = {\sigma^2_{gr}}/{\sigma^2_e}$ needs to be inferred from variance components estimated from individual records of the corresponding traits or other appropriate traits with individual records (Su *et al.*, 2018) or estimated using likelihood ratio test of the profile likelihood for different values of *b*.

Apart from the need of an optimal model for group records, accuracy of selection from the use of group records can increase with grouping strategies or allocation of animals into groups before phenotype testing. Accuracy of the prediction from group records increases when group members are more closely related based on pedigree information (Olson *et al.*, 2006; Peeters *et al.*, 2013; Su *et al.*, 2018). When genomic information is known before phenotype testing, utilization of genomic information can further increase accuracy of the prediction (Chapter 6). However, if all animals of each family are formed into a group, effects of direct additive genetics, maternal permanent environment and group cannot be separated. Here are some example designs of animal grouping to ensure separation of the direct additive genetic effect from maternal permanent environment and group effects in situations where group sizes are smaller, equal or greater than family sizes.

Family size > group size: Offspring from each family are used to form more than 1 group. Eg:

- Family size of 8 and group size of 4

- Design of grouping: Offspring from each family are allocated to two groups of 4 individuals.

Family size = group size: Each group has individuals from 2 different families. Eg:

- Family size of 8 and group size of 8
- Design of grouping: Group 1 consists of 4 individuals from family 1 and 4 individuals from family 2. Group 2 consists of 4 individuals from family 2 and 4 individuals from family 3 … Group n_g consists of 4 individuals from family n_g and 4 individuals from family 1.

Family size < group size: Offspring from more than one family of the same sire are used to form a group. Eg:

- Family size of 8 and group size of 16
- Design of grouping: Offspring from two families of the same sire are allocated to one groups.

In the examples above, hierarchical mating is assumed, meaning that one sire is mated to several dams, but one dam is mated to a single sire only. With the designs above, group members are closely related, and the maternal permanent environmental effect can be separated from the direct additive genetic effect with model M_{gr_exact} for group records. The use of genomic information for grouping, which was proposed in Chapter 6, can be applied for these designs to improve accuracy of selection.

The indirect additive genetic effect was attempted to be included in the model for group records (Peeters *et al.*, 2013). Through social interactions, birds can substantially affect one another's performance, and this effect is heritable (Ellen *et al.*, 2008; Peeters *et al.*, 2012). The direct additive genetic effect is the effect of an individual on its own performance, while the heritable effect of an individual on the performance of a group mate is known as the indirect additive genetic effect (Moore *et al.*, 1997; Willham, 1963; Wolf *et al.*, 1998). The indirect additive genetic effect can be estimated from individual records (Ellen *et al.*, 2008; Peeters *et al.*, 2012), but not from group records (Peeters *et al.*, 2013). Peeters *et al.* (2013) showed that the direct and indirect additive genetic effects cannot be separated in the model for group records. The incidence matrices that associate the direct and indirect genetic effects with phenotypic group records are identical. The covariance matrix for the direct additive genetic effect is proportional to the covariance matrix for the direct additive genetic effect, and thus only the total additive genetic effect can be estimated.

In summary, the use of group records is feasible to improve poultry breeding programs affected by GxE, especially when records from the production environments are only available in groups.

7.4. Benefits of genomic *versus* pedigree information for poultry breeding program in presence of GxE

Compared to the traditional pedigree-based breeding programs for dairy cattle, the main benefits of genomic information are the increase in accuracy of prediction, reduced generation intervals and reduced costs of progeny testing bulls (Hayes *et al.*, 2009). The increase in accuracy is due to a better measure of relationship between animals and better prediction of Mendelian sampling terms with genomic information whereas ability to predict the Mendelian sampling terms is missing with pedigree information in some cases. Those advantages may be smaller for poultry, particularly broilers, because poultry has short generation intervals, birds are usually phenotyped before sexual maturity and/or selection, and breeding programs for poultry often have a large number of selection candidates and high selection intensities (Wolc *et al.*, 2016). When candidates have own performances before selection, the Mendelian sampling terms are, to some extent, explored even with pedigree-based BLUP prediction. In addition, generation interval of broilers cannot be shortened with genomic information. However, when GxE interactions are strong, genomic information in GxE sib-testing schemes plays a major role in predicting the Mendelian sampling terms, and thus increases accuracy of selection. In GxE sib-testing schemes, selection candidates reside in B environment, and their sibs are tested for phenotypes in C environment. Because of bio-secure restrictions, those sibs in C environment are not used for selection, but only for provision of information on the desired performance for the selection candidates. Here are some theoretical benefits of genomic *versus* pedigree information for the "flow" of information from C to B environment:

- In a GxE sib test, the use of pedigree in a BLUP prediction exploits only up to 50% of the total genetic variance of the C trait. Selection candidates in B and birds in C can be full-sibs and half-sibs, and the highest relationship coefficient of birds between B and C is 0.5 (unrelated parents) using pedigree-based relationship matrix. Meanwhile, the use of genomic information in the prediction can exploit more than 50% of the total genetic variance of the C trait because the realized genomic relationships between specific full-sib pairs in broilers are varied. Observed ranges have been from 0.27 to 0.70 with mean of 0.47 and standard deviation of 0.05 (Lourenco *et al.*, 2015) or from 0.17 to 0.80 with mean of 0.48 and standard deviation of 0.09 (unreported results from the study of Chapter 2). Interestingly, the mean of full-sib relationships were less than 0.5 even the two broiler populations of Lourenco *et al.* (2015) and Chapter 2 were under intensive selection. Explanations from Lourenco *et al.* (2015) were genotyping errors, pedigree errors or loosely imposed quality control procedures. The use of pedigree in prediction of EBVs of selection candidates treats phenotypic performances of their full-sibs in C as an average

information. However, the phenotypic performances of their full-sibs in C are treated individually in genomic prediction of EBVs as the genomic relationships between specific full-sib pairs vary.

- In addition, information from related, more distantly related and even unrelated animals can be exploited in genomic prediction when genomic markers are in linkage disequilibrium with genotypes at casual loci (Daetwyler *et al.*, 2013). With pedigree-based prediction, information from unrelated individuals is not used for the prediction of EBVs of selection candidates. Genomic prediction utilizes linkage (more related to information from relatives) and linkage disequilibrium (more related to information from distant relatives and unrelated animals in population) (Daetwyler *et al.*, 2013). Therefore, accuracy of prediction depends on the density of genomic markers and the characteristics of the bird population in C (e.g. the size of dataset and the relatedness to selection candidates). Density of genomic markers regards to coverage of genomes, thus accuracy of prediction. For genomic prediction, the most common SNP chip panels have medium density (42k-60k) because their costs are relatively low and accuracy of prediction, as function of SNP density, seems to plateau at several tens of thousands of SNPs (Ilska *et al.*, 2014; Wolc *et al.*, 2016).

- Without selection candidates' own performances in C, Mendelian sampling terms for C traits are explored very little with pedigree information in a GxE sib-test, but they are explored better with genomic information. Ability to predict of the Mendelian sampling terms, which account for 50% of total genetic variance (no inbreeding), can lead to a much better prediction of EBVs with genomic information than with pedigree information.

When group records are used for selection, the genomic information can be utilized in two ways to improve accuracy of selection. One way is through the prediction model for group records that uses a genomic relationship matrix. Genomic information has a better measure of relationships between individuals, ability to predict Mendelian sampling terms and thus higher prediction accuracy of EBVs than pedigree information, just like for the prediction model of individual records. Genomic-based prediction from group records results in different EBVs of full-sibs that are in the same group. In contrast, the use of pedigree leads to identical EBVs for full-sibs in the same group. Therefore, genomic information can not only increase accuracy of EBVs, but reduce co-selection of full-sibs, and thus lower inbreeding compared to the use of pedigree information. The other way to utilize genomic information is animal grouping. Genotyping information was used to form groups by unsupervised and supervised clustering methods (Chapter 6). A slight additional accuracy was gained compared to the use of pedigree information for animal grouping. The clustering methods were particularly advantageous for accuracy of selection and rate of inbreeding when pedigree or dam information was missing.

Other benefit of genomic information for breeding programs affected by GxE regards rate of inbreeding. In Chapter 3, when proportion of birds moved to C increased from 15 to 45%, rate of inbreeding decreased. The decrease in the rate of inbreeding was due to a reduction in selection intensity and an increase in accuracy of prediction. The explanation of the increased accuracy was confirmed with an extra simulation in the discussion section of Chapter 3. The decrease in the rate of inbreeding due to the increased accuracy of prediction can only occur with genomic information, but not if only pedigree information is available. Genomic prediction reduces inbreeding by increasing the weight on Mendelian sampling terms whereas genomic prediction increases inbreeding by increasing the fixation rate of the favorable QTL allele (Pedersen *et al.*, 2010). The effect of increasing the weight on Mendelian sampling terms outperformed for inbreeding in the genomic selection breeding program of GxE sib-testing of Chapter 3.

As discussed in section 7.1, pedigree recording of crossbred chicken is usually missing, and thus performance records from crossbred birds in C cannot be used for EBV predictions for genetic evaluations of poultry breeding programs. However, this can be resolved with genotyping. Genomic information not only facilitates the "flow" of information from C to B environments, but accounts for GxG interactions due to crossbreeding.

7.5. Conclusions

Genotype-by-environment interactions (GxE) in poultry breeding programs and several approaches to improve genetic gains of the programs have been investigated. Environmental differences between a bio-secure breeding environment (B) and commercial production (C) conditions of broilers led to a strong GxE interaction for body weight (BW) traits: genetic correlations of 0.48-0.54 between BW traits measured in B and C environments, heterogeneous variances and different heritability for the B and C traits. In this thesis, investigated approaches to improve genetic gains of poultry breeding programs in the presence of GxE were: phenotyping strategies, optimal modelling of traits, use of group records, and use of genomic information. The key of these approaches is requirement of records obtained from C environments. The use of records in C can explore the re-ranking of EBVs for the two environments and unlock significant new sources of genetic variations as genetic variances of C traits can be more than 2 times higher than those of B traits.

Along with research studies carried out in this thesis, general discussion was extended to: GxE interactions for crossbred chicken raised in multiple commercial production environments, approaches for the development of statistical models, statistical models for group records, and theoretical benefits of

genomic information in case of GxE interactions. I believe that to improve poultry breeding programs affected by GxE, the use of records from purebred chicken in C is more reasonable than the use of records from crossbred chicken in C because the use of records from purebred in C has higher accuracy of selection, implementation of breeding schemes is less complicated, and chicken may have relatively low GxG interactions due to crossbreeding. It is also discussed that there are number of tools and approaches for evaluating statistical models. However, there are lack of comparisons between model evaluation methods, and it is little known which evaluation method is best for traits of interests in poultry breeding programs. In addition, the use of group records is feasible to improve poultry breeding programs affected by GxE, especially when records from the production environments are only available in groups. The use of genomic information can improve accuracy of selection, explore further variances of traits in C, increase potential benefits from group records, and reduce rate of inbreeding compared to the use of pedigree only.

7.6. Recommendations for future research

In this PhD thesis, the commercial production conditions included only one single environment. In Chapter 2, the C environment resembled commercial-like conditions. However, conditions of C environment can vary extensively for hygienic conditions, diseases, diets and management. Therefore, further research is needed to investigate GxE genetic parameters and breeding schemes for broilers when bird performances from multi-environments are recorded. Tracing pedigree and recording of birds from multiple commercial environments is difficult, but genotyping may ease the implementation.

Chapter 4 has proposed a cross-validation approach for the choice of statistical models when the maternal effects might be significant for the analyzed traits. The approach was based on theoretical assumptions of quantitative genetics, but the validity of the approach is unconfirmed. In addition, while optimal modelling of traits plays an important role for increasing accuracy of selection, methodologies to find the optimal model need extensive comparisons for validity of their use.

References:

Akaike, H 1973, 'Maximum likelihood identification of Gaussian autoregressive moving average models', *Biometrika,* vol. 60, no. 2, pp. 255-265.

Bijma, P 2012, 'Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations', *J Anim Breed Genet,* vol. 129, no. 5, pp. 345-58.

Biscarini, F, Bovenhuis, H, Ellen, ED, Addo, S & van Arendonk, JA 2010, 'Estimation of heritability and breeding values for early egg production in laying hens from pooled data', *Poult Sci,* vol. 89, no. 9, pp. 1842-9.

Biscarini, F, Bovenhuis, H & van Arendonk, JA 2008, 'Estimation of variance components and prediction of breeding values using pooled data', *Journal of Animal Science,* vol. 86, no. 11, pp. 2845-52.

Bolet, G, Garreau, H, Joly, T, Theau-Clement, M, Falieres, J, Hurtaud, J & Bodin, L 2007, 'Genetic homogenisation of birth weight in rabbits: Indirect selection response for uterine horn characteristics', *Livestock Science,* vol. 111, no. 1, pp. 28-32.

Christensen, OF, Madsen, P, Nielsen, B, Ostersen, T & Su, G 2012, 'Single-step methods for genomic evaluation in pigs', *Animal,* vol. 6, no. 10, pp. 1565-71.

Chu, TT, Alemu, SW, Norberg, E, Sørensen, AC, Henshall, J, Hawken, R & Jensen, J 2018, 'Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 52.

Daetwyler, HD, Calus, MP, Pong-Wong, R, de Los Campos, G & Hickey, JM 2013, 'Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking', *Genetics,* vol. 193, no. 2, pp. 347-65.

Duenk, P, Calus, MPL, Wientjes, YCJ, Breen, VP, Henshall, JM, Hawken, R & Bijma, P 2019, 'Estimating the purebred-crossbred genetic correlation of body weight in broiler chickens with pedigree or genomic relationships', *Genetics Selection Evolution,* vol. 51, no. 1, pp. 6.

Ellen, ED, Visscher, J, van Arendonk, JA & Bijma, P 2008, 'Survival of laying hens: genetic parameters for direct and associative effects in three purebred layer lines', *Poult Sci,* vol. 87, no. 2, pp. 233-9.

Godinho, RM, Bergsma, R, Silva, FF, Sevillano, CA, Knol, EF, Lopes, MS, Lopes, PS, Bastiaansen, JWM & Guimarães, SEF 2018, 'Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs', *Journal of Animal Science,* vol. 96, no. 3, pp. 817-829.

Gorjanc, G, Bijma, P & Hickey, JM 2015, 'Reliability of pedigree-based and genomic evaluations in selected populations', *Genetics Selection Evolution,* vol. 47, no. 1, pp. 65.

Grosso, JL, Balieiro, JC, Eler, JP, Ferraz, JB, Mattos, EC & Michelan Filho, T 2010, 'Comparison of different models to estimate genetic parameters for carcass traits in a commercial broiler line', *Genet Mol Res,* vol. 9, no. 2, pp. 908-18.

Habier, D, Götz, KU & Dempfle, L 2007, 'Estimation of genetic parameters on test stations using purebred and crossbred progeny of sires of the Bavarian Piétrain', *Livestock Science,* vol. 107, no. 2, pp. 142-151.

Hayes, BJ, Bowman, PJ, Chamberlain, AJ & Goddard, ME 2009, 'Invited review: genomic selection in dairy cattle: progress and challenges', *J Dairy Sci,* vol. 92, no.

Herrero-Medrano, JM, Mathur, PK, ten Napel, J, Rashidi, H, Alexandri, P, Knol, EF & Mulder, HA 2015, 'Estimation of genetic parameters and breeding values across challenged environments to select for robust pigs', *J Anim Sci,* vol. 93, no. 4, pp. 1494-502.

Hidalgo, AM, Bastiaansen, JWM, Groenen, MAM, Lopes, MS, Veroneze, R & de Koning, D-J 2015, 'Accuracy of genomic prediction using deregressed breeding values estimated from purebred and crossbred offspring phenotypes in pigs1', *Journal of Animal Science,* vol. 93, no. 7, pp. 3313-3321.

Ilska, J, Kranis, A & Woolliams, J 2014, 'The effect of training population size and chip density on accuracy and bias of genomic predictions in broiler chickens', 10th World Congress of Genetics Applied to Livestock Production, Vancouver, BC, pp.

Janhunen, M, Koskela, J, Ninh, NH, Vehviläinen, H, Koskinen, H, Nousiainen, A & Thỏa, NP 2016, 'Thermal sensitivity of growth indicates heritable variation in 1-year-old rainbow trout (Oncorhynchus mykiss)', *Genetics Selection Evolution,* vol. 48, no. 1, pp. 94.

Jasouri, M, Zamani, P & Alijani, S 2017, 'Dominance genetic and maternal effects for genetic evaluation of egg production traits in dual-purpose chickens', *Br Poult Sci,* vol. 58, no. 5, pp. 498-505.

Kapell, DN, Hill, WG, Neeteson, AM, McAdam, J, Koerhuis, AN & Avendano, S 2012, 'Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments', *Poultry Science,* vol. 91, no. 3, pp. 565-74.

Kheirabadi, K & Rashidi, A 2019, 'Modelling and genetic evaluation of Markhoz goat growth curve parameters', *Small Ruminant Research,* vol. 170, no. 43-50.

Knap, PW & Su, G 2008, 'Genotype by environment interaction for litter size in pigs as quantified by reaction norms analysis', *Animal,* vol. 2, no. 12, pp. 1742-7.

Legarra, A & Reverter, A 2017, 'Can we frame and understand cross-validation results in animal breeding?', Proc. Assoc. Advmt. Anim. Breed. Genet. 22, pp. 73-80.

Legarra, A & Reverter, A 2018, 'Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 53.

Legarra, A, Robert-Granié, C, Manfredi, E & Elsen, J-M 2008, 'Performance of Genomic Selection in Mice', *Genetics,* vol. 180, no. 1, pp. 611.

Lourenco, DAL, Fragomeni, BO, Tsuruta, S, Aguilar, I, Zumbach, B, Hawken, RJ, Legarra, A & Misztal, I 2015, 'Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken', *Genetics Selection Evolution,* vol. 47, no. 1, pp. 56.

Maniatis, G, Demiris, N, Kranis, A, Banos, G & Kominakis, A 2013, 'Model comparison and estimation of genetic parameters for body weight in commercial broilers', *Canadian journal of animal science,* vol. 93, no. 1, pp. 67-77.

Mäntysaari, E, Liu, Z & VanRaden, P 2010, 'Interbull validation test for genomic evaluations', *Interbull bulletin*, no. 41, pp. 17.

Mathur, PK, Herrero-Medrano, JM, Alexandri, P, Knol, EF, ten Napel, J, Rashidi, H & Mulder, HA 2014, 'Estimating challenge load due to disease outbreaks and other challenges using reproduction records of sows', *J Anim Sci,* vol. 92, no. 12, pp. 5374-81.

Mebratie, W 2019, 'The genetics of body weight and feed efficiency in broiler chickens, Aarhus University, Denmark.

Moore, AJ, Brodie III, ED & Wolf, JB 1997, 'Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions', *Evolution,* vol. 51, no. 5, pp. 1352-1362.

Mrode, RA & Thompson, R 2005, 'Linear models for the prediction of animal breeding values', CABI Publishing, Wallingford.

Mulder, H 2007, 'Methods to optimize livestock breeding programs with genotype by environment interaction and genetic heterogeneity of environmental variance, Wageningen University, The Netherlands.

Mulder, HA, Hill, WG, Vereijken, A & Veerkamp, RF 2009, 'Estimation of genetic variation in residual variance in female and male broiler chickens', *Animal,* vol. 3, no. 12, pp. 1673-80.

Myung, JI & Pitt, MA 2018, 'Model comparison in psychology', *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience,* vol. 5, no. 1-34.

Neves, HH, Carvalheiro, R & Queiroz, SA 2012, 'Genetic and environmental heterogeneity of residual variance of weight traits in Nellore beef cattle', *Genetics Selection Evolution,* vol. 44, no. 1, pp. 19.

Nielsen, B, Christensen, OF & Velander, I 2016, 'Nurse capacity, fertility, and litter size in crossbred sows and genetic correlation to purebred sow information', *J Anim Sci,* vol. 94, no. 5, pp. 1827-33.

Nurgiartiningsih, VM, Mielenz, N, Preisinger, R, Schmutz, M & Schueler, L 2004, 'Estimation of genetic parameters based on individual and group mean records in laying hens', *British Poultry Science,* vol. 45, no. 5, pp. 604-10.

Olson, KM, Garrick, DJ & Enns, RM 2006, 'Predicting breeding values and accuracies from group in comparison to individual observations', *Journal of Animal Science,* vol. 84, no. 1, pp. 88-92.

Pedersen, L, Sørensen, A & Berg, P 2010, 'Marker-assisted selection reduces expected inbreeding but can result in large effects of hitchhiking', *Journal of Animal Breeding and Genetics,* vol. 127, no. 3, pp. 189-198.

Peeters, K, Ellen, ED & Bijma, P 2013, 'Using pooled data to estimate variance components and breeding values for traits affected by social interactions', *Genetics Selection Evolution,* vol. 45, no.

Peeters, K, Eppink, TT, Ellen, ED, 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,* vol. 192, no. 2, pp. 705-713.

Posht-e Masari, HA, Hafezian, SH, Abdollahi-Arpanahi, R, Mokhtari, MS, Mianji, GR & Yeganeh, AT 2019, 'The comparison of alternative models for genetic evaluation of growth traits in Lori-Bakhtiari sheep: Implications on predictive ability and ranking of animals', *Small Ruminant Research,* vol. 173, no. 59-64.

Putz, AM, Tiezzi, F, Maltecca, C, Gray, KA & Knauer, MT 2018, 'A comparison of accuracy validation methods for genomic and pedigree-based predictions of swine litter size traits using Large White and simulated data', *J Anim Breed Genet,* vol. 135, no. 1, pp. 5-13.

Rönnegård, L, Felleki, M, Fikse, WF, Mulder, HA & Strandberg, E 2013, 'Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle', *Journal of Dairy Science,* vol. 96, no. 4, pp. 2627-2636.

SanCristobal-Gaudy, M, Elsen, J-M, Bodin, L & Chevalet, C 1998, 'Prediction of the response to a selection for canalisation of a continuous trait in animal breeding', *Genetics, Selection, Evolution: GSE,* vol. 30, no. 5, pp. 423-451.

Santana, ML, Jr., Bignardi, AB, Eler, JP & Ferraz, JB 2016, 'Genetic variation of the weaning weight of beef cattle as a function of accumulated heat stress', *J Anim Breed Genet,* vol. 133, no. 2, pp. 92-104.

Sartori, C, Tiezzi, F, Battagin, M, Guzzo, N & Mantovani, R 2018, 'Genotype by environment interactions in productive traits in a local cattle breed due to breeding area, farming systems and feeding strategies'. "*Proceedings of the World Congress on Genetics Applied to Livestock Production*", vol. Electronic Poster Session - Challenges - Genotype by Environment Interactions, pp.918

Schwarz, G 1978, 'Estimating the dimension of a model', *The annals of statistics,* vol. 6, no. 2, pp. 461-464.

Silva, FF, Mulder, HA, Knol, EF, Lopes, MS, Guimaraes, SE, Lopes, PS, Mathur, PK, Viana, JM & Bastiaansen, JW 2014, 'Sire evaluation for total number born in pigs using a genomic reaction norms approach', *J Anim Sci,* vol. 92, no. 9, pp. 3825-34.

Spiegelhalter, DJ, Best, NG, Carlin, BP & Van Der Linde, A 2002, 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 64, no. 4, pp. 583-639.

Su, G, Guldbrandtsen, B, Gregersen, VR & Lund, MS 2010, 'Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population'. *In* '*J Dairy Sci*'.

Su, G, Madsen, P, Lund, MS, Sorensen, D, Korsgaard, IR & Jensen, J 2006, 'Bayesian analysis of the linear reaction norm model with unknown covariates', *J Anim Sci,* vol. 84, no. 7, pp. 1651-7.

Su, G, Madsen, P, Nielsen, B, Ostersen, T, Shirali, M, Jensen, J & Christensen, OF 2018, 'Estimation of variance components and prediction of breeding values based on group records from varying group sizes', *Genetics Selection Evolution,* vol. 50, no. 1, pp. 42.

Van Vleck, LD 1993, 'Variance of prediction error with mixed model equations when relationships are ignored', *Theor Appl Genet,* vol. 85, no. 5, pp. 545-9.

Wei, M & Werf, JHJ 1995, 'Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits', *J Anim Sci,* vol. 73, no.

Wientjes, YCJ & Calus, MPL 2017, 'BOARD INVITED REVIEW: The purebred-crossbred correlation in pigs: A review of theory, estimates, and implications1', *Journal of Animal Science,* vol. 95, no. 8, pp. 3467-3478.

Willham, RL 1963, 'The covariance between relatives for characters composed of components contributed by related individuals', *Biometrics,* vol. 19, no.

Wolc, A, Kranis, A, Arango, J, Settar, P, Fulton, JE, O'Sullivan, NP, Avendano, A, Watson, KA, Hickey, JM, de los Campos, G, Fernando, RL, Garrick, DJ & Dekkers, JCM 2016, 'Implementation of genomic selection in the poultry industry', *Animal Frontiers,* vol. 6, no. 1, pp. 23-31.

Wolc, A, White, IM, Avendano, S & Hill, WG 2009, 'Genetic variability in residual variation of body weight and conformation scores in broiler chickens', *Poult Sci,* vol. 88, no. 6, pp. 1156-61.

Wolf, JB, Brodie III, ED, Cheverud, JM, Moore, AJ & Wade, MJ 1998, 'Evolutionary consequences of indirect genetic effects', *Trends in ecology & evolution,* vol. 13, no. 2, pp. 64-69.

Wondmeneh, E, Van der Waaij, EH, Udo, HMJ, Dessie, T & Van Arendonk, JAM 2016, 'Comparison of different poultry breeds under station and on-farm conditions in Ethiopia', *Livestock Science,* vol. 183, no. 72-77.

Xiang, T, Christensen, OF, Vitezica, ZG & Legarra, A 2016, 'Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs', *Genet Sel Evol,* vol. 48, no. 1, pp. 92.

Zerehdaran, S, Vereijken, Aj, Arendonk, Jv & Van der Waaij, E 2005, 'Effect of age and housing system on genetic parameters for broiler carcass traits', *Poultry science,* vol. 84, no. 6, pp. 833-838.

Zumbach, B, Misztal, I, Azain, M, Tsuruta, S, Sanchez, JP, Holl, J, Long, T, Herring, W & Culbertson, M 2008, 'Genetic components of heat stress in finishing pigs: Parameter estimation', *Journal of Animal Science,* vol. 86, no. 9, pp. 2076-2081.

Summary
&
Sammendrag

Within production of commercial broilers, differences between the breeding (B) and commercial production (C) environments may be related to e.g. differences in hygienic conditions, climate and management. For indigenous chicken, the differences may be also related to availability of feed, sub-optimal diets, predators and social interactions of birds with other livestock species. These environmental differences may lead to genotype by environment (GxE) interactions i.e. a re-ranking of genotypes between B and C environments is expected. A substantial re-ranking implies that genetic progress achieved in the breeding programs is not fully realized in production animals' performances. This PhD-project investigated GxE interactions in breeding programs for poultry and presented solutions to improve genetic progress of these breeding programs.

In Chapter 2, GxE for body weight (BW) was explored for broilers raised in B and C environments. A strong GxE interaction was found, as significant re-ranking, heterogeneous variances and different heritability for B and C traits were found. The genetic correlations between BW traits measured in B and C environments were in the range 0.48-0.54. Genetic variances of C traits were more than 2 times higher than those of B traits. Heritability of C traits (0.31-0.37) were higher than those of B traits (0.27-0.30). In addition, genomic information was used to increase accuracy of prediction in the presence of GxE interactions. It was found that the correlation between EBVs of C traits estimated from the full dataset and reduced dataset increased by 31-37% for genotyped validation birds and 6-15% for non-genotyped validation birds when the use of combined pedigree and genomic information was compared to the use of pedigree information only. The use of combined pedigree and genomic information reduced bias of prediction for the genotyped birds, but increased bias of prediction for the non-genotyped birds.

In Chapter 3, breeding schemes with different proportions of birds that were phenotyped in B and C environments were compared in a genomic selection breeding program for broilers. It was found that when the genetic correlations between traits measured in B and C were 0.5 and 0.7, allocating 70% and 30% hatched birds to B and C environments, respectively, for phenotype testing led to the highest genetic gains among the compared phenotyping strategies. When the genetic correlation was 0.9, birds moved to C did not improve genetic gains of the breeding scheme due to reduced selection intensity. Increasing proportion of birds moved to C (from 15 to 45%) could reduce rate of inbreeding of the breeding program.

In Chapter 4, a genetic analysis was carried out for BW in broilers at different ages raised in a commercial environment. A statistical model was developed with the aim to increase predictive ability of the model for the traits affected by maternal effects. A criterion in the development of the statistical model was based on correlation between EBVs and corrected phenotypes of half-sib individuals. The statistical model also accounted for heterogeneous variances between sexes. It was found that variance components for

BW increased drastically with increasing age of broilers. The drastic increase in variances were mainly due to scaling effects. The difference in variance components between sexes increased with increasing age. The permanent environmental maternal effect reduced gradually as broilers aged.

In Chapter 5, the use of group records was investigated in breeding programs for village chicken, of which strong GxE interactions are expected due to large difference between breeding station and the village environments. This was modelled for two traits, growth and egg production. It was found that the use of group records from village significantly improved genetic gains compared to the scheme without birds tested in the village although group records led to a slightly lower genetic gain compared to individual records.

In Chapter 6, genomic information was utilized to form groups before phenotyping. Two grouping methods based on genomic information were proposed: unsupervised clustering implemented in the STRUCTURE software and supervised clustering based on genomic relationships. Compared to the grouping methods based on pedigree information only, the grouping methods based on genomic information resulted in higher relationships between group members, and thus increased accuracy of GEBV prediction (1.2-11.7%).

Chapter 7 (General discussion) extends the discussions of GxE interactions for crossbred animals, GxE interactions in multiple commercial production environments and GxE interactions due to micro-environmental differences. Tools for the development and comparison of statistical models are discussed including model goodness-of-fit, accuracy of EBVs and cross-validation approaches. In addition, different prediction models for group records are compared when the group effect and permanent environmental maternal effect are accounted for in the models. Theoretical benefits of genomic information for breeding programs affected by GxE are discussed.

To conclude, differences between the breeding and production environments can lead to substantial GxE interactions. In the presence of GxE interactions, a breeding program for poultry should establish recording systems under the production environments in either individual or group records in order to ensure maximum genetic gains and provide customers with genotypes well adapted to the production environments. In addition, an optimal cross-validation procedure for better choice of statistical models is needed for genetic evaluations in poultry breeding programs as better modelling of traits is a low-cost approach to improve accuracy of selection.

Indenfor kommerciel slagtekyllingproduktion kan forskelle mellem avls- (B) og kommercielle produktionsmiljøer (C) være relateret til forskelle i blandt andet hygiejne klima, management. For de oprindelige racer kan forskellene også relateres til tilgængelighed af foder, suboptimal foder, rovdyr og sociale interaktioner med andre husdyrarter. Disse miljøforskelle kan føre til genotype-miljø (GxE) interaktioner, dvs. at der forventes en omrangering af genotyper mellem B- og C-miljøer. En betydelig omrangering indebærer, at genetiske fremskridt opnået i avlsprogrammerne, ikke fuldt ud vil realiseres i det kommercielle produktionsmiljø. Dette ph.d.-projekt undersøgte GxE-interaktioner i avlsprogrammer for fjerkræ samt løsninger til forbedring af genetisk fremgang for disse avlsprogrammer.

I kapitel 2 blev GxE for kropsvægt (BW) undersøgt for slagtekyllinger opdrættet i B- og C-miljøer. Der blev fundet en stærk GxE-interaktion, med markant omrangering av dyr, heterogen varians og forskellige arvbarheder for egenskaber målt i B- og C miljøet. De genetiske korrelationer mellem BW-egenskaber målt i B- og C-miljø lå i området 0,48-0,54. Genetisk varians for BW målt i C-miljøet var mere end 2 gange så høj som for BW målt i B-miljøet. Arvbarheden af egenskaben i C-miljøet (0,31-0,37) var højere end for B-miljøet (0,27-0,30). Derudover blev genomisk information brugt til at øge sikkerheden på de predikerede EBV'er under tilstedeværelse av GxE-interaktioner. Det blev fundet, at sammenhængen mellem EBV'er estimeret ut fra det fulde og det reducerede datasæt steg med 31-37% for genotypede fugle og 6-15% for ikke-genotype fugle, når brugen af kombineret stamtavle og genomisk information var sammenlignet med brugen af kun stamtavleinformation. Anvendelsen af kombineret stamtavle og genomisk information reducerede prediktions bias for de genotypede fugle, men øgede den for de ikke-genotypede fugle.

I kapitel 3 blev avlsplaner med forskellige andele af fugle, der blev fænotypisk testet i B- og C-miljø, sammenlignet i et avlsprogram med genomisk selektion. Det blev fundet, at når de genetiske sammenhænge mellem egenskaber målt i B og C var 0,5 og 0,7, førte allokering af henholdsvis 70% og 30% af fænotyper på fugle i miljø B- og C, til den højeste genetiske fremgang blandt de sammenlignede fænotype strategier. Når den genetiske korrelation var 0,9, resulterede en øget fænotypning af fugle i miljø C ikke til en øgning i genetisk fremgang på grund af lavere selestionsintensitet. At øge andel af fugle, der blev fænotyper i C (fra 15 til 45%), kunne reducere indavlsraten af avlsprogrammet.

I kapitel 4 blev der udført en genetisk analyse af BW ved forskellige aldre på slagtekyllinger opvokset i et kommercielt miljø. En statistisk model blev udviklet med det formål at øge modelens prediktionsevne for de egenskaber, der var påvirket af maternale effekter. Modellen var baseret på korrelation mellem EBV'er og korrigerede fænotyper af individer med halvsøskende. Den statistiske model tog også hensyn

til heterogen varians mellem kønnene. Det blev fundet, at varianskomponenter for BW steg drastisk med stigende alder på slagtekyllinger. Dette skyldtes hovedsageligt skaleringseffekter. Forskellen i varianskomponenter mellem kønnene steg med stigende alder. Den permanente miljømæssige effekt af mødrene reduceres gradvist, når slagtekyllinger ældes.

I kapitel 5 undersøges anvendelsen af grupperegistreringer i avlsprogrammer til landsbykylling, hvor der forventes en stærk GxE-interaktion pga store forskelle i miljøet mellem test/avlsstationerne og landsbyerne. Der blev simuleret egenskaber tilvækst og ægproduktion, og resultaterne viser, at brugen af grupperegistreringer fra landsbyen, signifikant forbedrede genetisk fremgang sammenlignet med en strategi hvor fugle kun blev testet på teststation. Brug af grupperegistreringer førte til en lidt lavere genetisk fremgang sammenlignet med brug af individuelle registreringer.

I kapitel 6 blev genomisk information brugt til at danne grupper inden registrering af fænotyper på gruppeniveau. To grupperingsmetoder baseret på genomisk information blev anvendt: unsupervised clustering implementeret i STRUCTURE-programmet og supervised clustering baseret på genomisk slægtskab. Sammenlignet med grupperingsmetoderne, der kun var baseret på stamtavleinformation, resulterede grupperingsmetoderne baseret på genomisk information i øget slægtskab mellem gruppemedlemmer og dermed øget sikkerhed på prediktionen af GEBV (1,2-11.7%).

I kapitel 7 (Generel diskussion) tages diskussionerne om GxE-interaktioner for krydsningsdyr, GxE-interaktioner i flere kommercielle produktionsmiljøer og GxE-interaktioner på grund af mikro-miljøforskelle videre. Værktøjer til udvikling og sammenligning af statistiske modeller diskuteres, inklusive goodness-of-fit, sikkerhederne af EBV'er og krydsvalideringsmetoder. Derudover sammenlignes forskellige prediktionsmodeller for grupperegistreringer. De teoretiske fordele ved bruk af genomisk information i avlsprogrammer med GxE diskuteres.

Det konkluderes, at forskelle mellem avls- og produktionsmiljøer kan føre til betydelige GxE-interaktioner. Hvis dette er tilfældet, bør det etableres registreringer i produktionsmiljøerne, enten på individuelle dyr eller i grupper, så en maksimal genetisk fremgang og genotyper, der er godt tilpasset produktionsmiljøerne, sikres. Derudover er der behov for en optimal krydsvalideringsprocedure for at vælge statistiske modeller til genetisk evaluering i fjerkræavlsprogrammer, da en bedre modellering af egenskaberne er en kost effektiv måde at forbedre sikkerheden på i selektionen.

Acknowledgements

Acknowledgements

My PhD study would be impossible to be ended without great support from many people to whom I own a million thanks. I take this opportunity to express, in words, a small part of my thankful feelings to the people who have eased the challenging time of my life.

The most influential person in my PhD study, particularly at the start of my PhD time, was Elise Norberg, my main supervisor at AU. You have been very supportive to me for both personal and study-related matters. In my current place, there are still stuff that you helped me to settle down in Viborg. With cattle nutrition background, I felt I was in the wrong office. Your teaching approach for me to catch up with the field of quantitative genetics was so soft and patient. I don't list all things you have done to me here, but I do remember them all. Thank you very much, Elise.

I would definitely not get a PhD degree without you, Just Jensen, my co-supervisor at AU. I did not believe when I first heard the legendary story of your scientific career. But shortly, I realized that you are simply amazing. All scientific knowledge I learnt during PhD was from you or derived from your encouragement, ugly-but-valuable board drawings and words-run-out-of-space comments. Thank you for teaching me, inspiring my scientific curiosity, and being a very thorough supervisor, Just.

I am very thankful to John Bastiaansen, my daily supervisor at WUR. Thank you for your kindness, listening skills and patience. You gave me very supportive consultation and intellectual freedom in my work. You made me feel confident about what I am capable of. I am very thankful to Hans Komen, Peer Berg and Per Madsen for your scientific advices and insightful suggestions. I appreciate very much Christian Sørensen for your help on my simulation studies. I also benefit a lot from courses by Jan Lassen (Linear models in animal breeding) and Bernt Guldbrandtsen (Quantitative genetics).

I would like to thank researchers from Cobb-Vantress for your data and valuable discussions on commercial breeding programs. I appreciate very much all co-authors for your contributions on my papers. A big thank goes to Setegn Alemu for the help on personal life and work. Thank to your recommendation, I started learning and loving R, my first programming language. I would like to thank Helene, Wossenie and number of officemates from both AU and WUR for sharing room, daily-life stories, good jokes and sometimes food.

I would like to acknowledge all QGG and ex-QGG people for helping me with administrative stuff, answering my stupid questions, sometimes, and creating a nice, warm and friendly, but professional environment for people to work in. I feel very fortunate to know friends from EGS-ABG groups like Mesbah, Grum, Juan and Vinicius. Did we give each other a hard time? I did not feel that from you, cheers.

I get not only support, but encouragement from colleagues, especially Do D Luc and Nguyen C Thanh, at Vietnam National University of Agriculture. It is very nice to meet some of you, Duong TT Huyen, Nguyen T Huyen, Nguyen H Thinh and Nguyen TC Giang, in Europe.

Last but not least, my PhD study would be impossible without full support from my family. Cháu cảm ơn bà đã yêu thương và luôn phù hộ cháu. Con cảm ơn bố, mẹ đã nuôi dạy con khôn lớn. Cảm ơn bố đã nghiêm khắc dạy dỗ, đã định hướng và giúp đỡ con rất nhiều trong công việc và học tập. Để đạt được bằng Tiến sỹ này, công lao của bố là rất lớn. Cảm ơn mẹ đã luôn động viên, và là cầu nối tình cảm của cả gia đình mình. Em/anh/cháu cảm ơn chị Thắm, em Vượng, bà con họ hàng nội ngoại đã luôn động viên tinh thần. Anh cảm ơn Hoài Anh đã đồng hành cùng anh trong suốt chương trình nghiên cứu sinh. Tuấn, Ngân yêu dấu, các con là niềm hạnh phúc mỗi khi bố về nhà, chăm ngoan và học giỏi nhé, hãy để bố tự hào về các con.

Curriculum Vitae

About the author

Thinh Tuan Chu was born in September 1984 Hung Yen, Vietnam. He started his tertiary education at Vietnam National University of Agriculture (VNUA), Vietnam, but then left to University of Queensland (UQ), Australia where he completed his Bachelor of Agricultural Science and Master of Animal Studies. His studies at UQ were covered by Australian Development Scholarship from AusAID - Australian Agency for International Development, and Scholarship of Biotechnology in Agriculture and Aquaculture from Vietnamese Ministry of Agriculture and Rural Development. The thesis topics of his bachelor and master's studies were isolation of bacteria from ruminal fluids, and comparison of methodologies for the digestion of forages, respectively. Thinh was employed at VNUA as a lecturer. His main teaching subjects at VNUA were animal genetics and breeding, and principles of quantitative genetics for undergraduate students. He also involved in research duties such as studies on animal production, conservation of local breeds and some molecular laboratory works. In 2015, Thinh received the Erasmus Mundus scholarship to pursue the joint doctorate program, EGS-ABG, at Aarhus University, Denmark and Wageningen University, Netherlands. After completion of the PhD study, Thinh will continue his scientific career as a postdoctoral researcher at Center for Quantitative Genetics and Genomics, Aarhus University.

Peer review publications

1. **Chu, T.T.**, Bastiaansen J.W.M., Berg P., Rome H.J.S., D. Marois, Henshall J., Jensen J., 2019. Use of genomic information to exploit genotype by environment interactions for body weight of broiler chicken in bio-secure and production environments. Genetics Selection Evolution, vol. 51:50. doi:10.1186/s12711-019-0493-3
2. **Chu, T.T.**, Bastiaansen J.W.M., Berg P., Komen H., 2019. Optimized grouping to increase accuracy of prediction of breeding values based on group records in genomic selection breeding programs. GSE accepted. DOI: 10.1186/s12711-019-0509-z
3. **Chu T.T.**, Madsen P., Norberg E., Wang L., Marois D., Henshall J., Jensen J., 2019. Genetic analysis on body weight at different ages in broiler chicken raised in commercial environment. Journal of Animal Breeding and Genetics: Accepted. doi:10.1111/jbg.12448
4. **Chu T.T.**, Alemu S.W., Norberg E., Sørensen A.C., Henshall J., Hawken R., Jensen J. (2018). 'Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken'. Genetics Selection Evolution, vol. 50:52. doi:10.1186/s12711-018-0430-x
5. **Chu T.T.**, Bastiaansen J.W.M., Norberg E., Berg P. (2018). On farm observations to increase genetic gain in breeding schemes for village poultry production – A simulation study. Acta Agriculturae Scandinavica, Section A — Animal Science, 1-10. doi:10.1080/09064702.2018.1543444.

Conference proceedings, abstracts and presentations

1. **Chu T.T.**, Bastiaansen J.W.M., Berg P., Komen H. (2019). Use of genomic information to improve accuracy of prediction from group records. In proceedings of the European Federation for Animal Science Annual Meeting (EAAP) 2019, Ghent, Belgium.
2. **Chu T.T.**, Norberg E., Huang C., Henshall J., Jensen J. (2019). Benefits of using genomic information for broiler breeding program in presence of GxE interactions. In proceedings of the European Federation for Animal Science Annual Meeting (EAAP) 2019, Ghent, Belgium.
3. **Chu T.T.**, Rome H.J.S., Norberg E., Marois D., Henshall J., Jensen J. (2019). GxE interactions of body weight for broilers raised in bio-secure and commercial environments. In proceedings of the European Federation for Animal Science Annual Meeting (EAAP) 2019, Ghent, Belgium.
4. Rome H.J.S., **Chu T.T.**, R. Hawken, J. Henshall, Jensen J. (2019). WGBLUP model improves accuracy of breeding values prediction in a commercial line of broilers. In proceedings of the European Federation for Animal Science Annual Meeting (EAAP) 2019, Ghent, Belgium.
5. **Chu T.T.**, Norberg E., Bastiaansen J.W.M., Berg P. (2018). On farm observations to increase genetic gain in breeding schemes for village poultry production. In proceedings of the 11th World Congress on Genetics Applied to Livestock Production (WCGALP), Auckland, New Zealand.
6. **Chu T.T.**, P. Madsen, L. Wang, J. Henshall, R. Hawken, J. Jensen (2018). Influence of age on variance components for body weight in commercial male and female broiler chicken. In proceedings of the European Federation for Animal Science Annual Meeting (EAAP) 2018, Dubrovnik, Croatia.
7. **Chu T.T.**, S.W. Alemu, E. Norberg, A.C. Sørensen, J. Henshall, and J. Jensen. 2017. Benefits of testing birds in both bio-secure and production environment in genomic selection breeding programs for commercial broiler chicken. 10th European Symposium on Poultry Genetics (ESPG), St. Malo, France.

Individual Training Plan (ITP)

Training (36.5 ECTS)		
Mandatory courses (7.5)	**Place**	**Year**
Welcome to EGS-ABG	SLU, Sweden	2015
EGS-ABG Summer Research School: The sustainability concept in animal breeding	SLU, Sweden	2015
EGS-ABG Summer Research School: Emerging technology	WUR	2017
Scientific Integrity & Ethics and Animal Science	WUR	2019
Advanced scientific courses (22 ECTS)		
Quantitative genetics	AU	2015
Linear models in animal breeding	AU	2016
Feed efficiency in dairy cattle	AU	2016
Animal breeding planning	AU	2016
Quantitative genetics in animal breeding (NOVA)	UH, Finland	2016
Professional Skills support courses (7 ECTS)		
Academic English for non-Danish speaking PhD students	AU	2018
Techniques for Writing and Presenting a Scientific Paper	AU	2017
Design and implementation of breeding programs for smallholder poultry farmers	ILRI, Ethiopia	2015
Knowledge dissemination		
Teaching		
Co-supervision of BSc student	AU	2016
International conferences		
70th European Annual Meeting of the European Federation of Animal Science (EAAP2019)	Belgium	2019
69th European Annual Meeting of the European Federation of Animal Science (EAAP2018)	Croatia	2018
10th European Symposium on Poultry Genetics	France	2017
Seminars and workshop		
MBG- AU PhD conference 2016	AU	2016
MBG- AU PhD conference 2017	AU	2017

Colophon

The PhD study was supported from a joint grant from the European commission, within the framework of the Erasmus-Mundus joint doctorate program (EGS-ABG), and Graduate School of Science and Technology, Aarhus University. The studies used the stochastic simulation ADAM from QGG, and data from Cobb-Vantress, Siloam Springs, Arkansas 72761-1030, United States of America.

This thesis was printed by Digisource, Viborg, Denmark