Predicting breed by environment interaction using ecological modelling

María Lozano Jaramillo

Propositions

1. Ecological models can accurately predict breed environmental suitability.
   (this thesis)

2. Environmental sensitivity is underestimated when testing breeds in classical breed by country experiments.
   (this thesis)

3. Current research focuses on how livestock contributes to climate change, but neglects the impact of climate change on livestock.

4. While living in a different culture, integration is a process under natural selection.

5. Academic education should not be restricted to one institution only.

6. Humanity should not mobilize to restore, but to leave alone what we still have not lost.

Propositions belonging to the thesis entitled:

Predicting breed by environment interaction using ecological modelling

María Lozano-Jaramillo

Wageningen, 1 November 2019

Predicting breed by environment interaction using ecological modelling

María Lozano-Jaramillo

Thesis committee

Promotor
Prof. Dr J Komen
Personal chair at Animal Breeding and Genomics
Wageningen University & Research

Co-promotor
Dr JWM Bastiaansen
Researcher, Animal Breeding and Genomics
Wageningen University & Research

Prof. Dr T Dessie
Senior Scientist, Animal Breeding and Genetics
International Livestock Research Institute

Other members
Prof. Dr AK Bregt, Wageningen University & Research
Prof. A Djikeng, University of Edinburgh, The Roslin Institute, UK
Prof. Dr J Sölkner, BOKU University, Vienna, Austria
Prof. Dr BJ Zwaan, Wageningen University and Research

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS).

Predicting breed by environment interaction using ecological modelling

María Lozano-Jaramillo

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday November 1, 2019
at 4 p.m. in the Aula.

Lozano-Jaramillo, M.
Predicting breed by environment interaction using ecological modelling
150 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references, with summary in English

ISBN 978-94-6395-071-8
DOI: https://doi.org/10.18174/497758

Abstract

Lozano-Jaramillo, M. (2019). Predicting breed by environment interaction using ecological modelling. PhD thesis, Wageningen University, the Netherlands.

Animal production plays an important socio-economic role in developing countries. Livestock largely contribute to the local food production, and they represent a source of income for many smallholder farmers. Smallholder production systems are characterized by low producing indigenous breeds. To enhance food security, several breeding projects have introduced commercial breeds as an alternative to help smallholder farmers to increase the food supply. As a common practice in animal breeding, commercial breeds are introduced from a nucleus to different locations. However, no knowledge on the new environmental conditions in the new tropical environment are taken into account. The aim of this thesis was to understand how the environment plays a role in shaping the variation in breed performance across different agro-ecologies. To achieve this, I used different approaches than what it is commonly used in animal breeding, to understand the environmental variation. I use Geographic Information Systems (GIS) along with different models used in ecology to predict the performance of genotypes in changing environments. Results highlight the importance of acknowledging the environment being composed of many variables that have continuous variation when designing breeding programs or testing schemes. I recommend the application of these models to livestock to make predictions on how productivity traits respond as a function of the environment.

Contents

Chapter 1

General introduction

1.1 Genotype by environment interaction

The aim in animal breeding is to select and breed animals that will perform efficiently in future production settings. How well these animals perform depends on the genotype and the production environments where they are kept. Genotype by environment interaction (G x E) occurs when genotypes respond differently to changes in the environment. The presence of G x E can be determined by estimating the genetic correlation (r_g) between the measurements for a given trait in two environments. Selective breeding requires several conditions such as large populations, accurate pedigree and production recording, as well as a reduction of the environmental variation. Conditions that make breeding programs challenging to accomplish in developing countries (Besbes, 2009; Dana et al 2010). A way to overcome these obstacles is by developing the breeding program on a station or other controlled system. Then a new problem arises; the overall management conditions between on station and on farm are different, which may lead to a G x E (Lwelamira, 2012). It is important to determine the magnitude of the G x E if there is a significant re-ranking in the performance of a breed between the two environments, and to establish if different breeding programs are needed to account for this (Falconer, 1952; Mulder and Bijma, 2005).

When evaluating the presence of G x E, the environment is normally represented as different geographical locations or management systems or a combination of both. Generally, no information on the specific environmental factors that define the variability between the different locations is acknowledged. In some cases, when commercial breeds are exposed to tropical conditions, a decrease in productivity is observed as a result of the differences between the central breeding station and the smallholder farm, but there is no knowledge of what environmental factors are causing this change in performance. To establish which breeds should be used to increase local meat and egg production, there should be a better understanding of how the different environmental characteristics play a role in the growth and survival of these breeds. Having this knowledge can shed light on how the breeds will locally adapt to harsh environments, and may allow the optimization of livestock production by finding breeds that can withstand specific, or a wide range of environmental settings.

1.2 Distribution modelling

A main focus in ecology has been to understand the relationship between a species and its environment. In other words, to explain patterns of animal and plant distributions as a result of climate and other environmental factors. Geographic information systems (GIS) are computer-based tools that have become of importance in ecological research as they support in handling, collecting, and analyzing geographic or spatial data (Guisan and Zimmermann, 2000). The application of GIS tools has increased over the years, not only in ecology, but also in other disciplines such as conservation, geography and environmental science. However, in livestock research, the use of these tools is still limited (Joost et al., 2010).

Predictive habitat distribution models are GIS-based tools that use location data (e.g geo-references) and relate these observations to environmental variables to make predictions of potential geographic ranges (Guisan and Zimmermann, 2000; Anderson et al., 2002). To elucidate these relationships between environmental variables, ecologists often use regression analyses. However, because the ecological data is complex, showing non-linear relationships, or spatial or temporal correlations, different modelling approaches have been used to address these issues (Maloney et al., 2012). Machine learning algorithms have been used to improve the prediction accuracy of the ecological models. Maximum entropy algorithm, implemented by Maxent (Phillips et al., 2006), is one of the most commonly used tools in ecology that relates environmental data and occurrence locations to establish the probability of potential geographical suitability (Phillips and Dudík, 2008). It is commonly used because the algorithm uses presence-only data, performs variable selection to avoid overfitting, and it has been shown to have more predictive power than other algorithms, particularly for small data sets (Elith et al., 2006).

A different approach that has been recently introduced, are models that correlate phenotypes to environmental variables to map the phenotypes within the species distribution. These phenotypic distribution models can capture the response of a phenotypic trait as a function of the environmental conditions. Different algorithms, such as gradient boosting, have shown to improve the prediction accuracy of these ecological models over using standard regression models (Maloney et al., 2012). They have been applied in studies relating phenotypic and environmental variation in wild species. In a species of stream fish, these models were used to predict future phenotypes as a result of future changes in stream flow rates (Michel et al., 2017). A different study used the phenotypic distribution models to predict the changes in phenotypes of a dominant prairie grass as a response to different climate change scenarios (Smith et al., 2017). These models have not yet been applied in livestock research.

The use of GIS tools, and ecological models can therefore be an innovative tool to apply to livestock. These models can be used to explain differences or similarities in productivity between breeds. They can also be used to make predictions of which areas can be suitable to introduce a breed. Most importantly, they can be used in properly characterizing the environment when evaluating the existence of a mismatch in performance in different environments.

Glossary
Geographic information systems (GIS) Computer system for gathering, managing, analyzing, and/or presenting geographic or spatial data.
Gradient boosting Machine learning technique for regression fitting that produces a prediction model from an initial estimation of a function that fits the data poorly. The fit is improved in each successive iteration.
Maxent Machine learning algorithm that uses presence-only data to relate the location records of a species and the environmental characteristics to establish a probability of potential suitability.
Phenotypic distribution models Use associations between phenotypic traits and environmental variables to map the phenotypic traits of populations within that species' distribution.
Predictive habitat distribution models Use information on the current environmental characteristics for a species under a particular set of location records, to generate predictions of areas of potential habitat suitability for the species.

1.3 Breeding for different agro-ecologies

In natural populations, animals are exposed to natural conditions that can have an effect on how their phenotypes are expressed. When a population is distributed across different environments, individuals are exposed to different selective pressures (Sexton et al., 2009). This will promote local adaptation in different populations, such that individuals will exhibit better performance in their native environment relative to others (Joshi et al., 2001). In livestock, adaptation to the environment can be critical, especially in developing countries. The most productive animals must be identified for breeding purposes, as the sign that they are locally adapted is reflected in their performance. Failure to adapt can lead to loss in productivity, disease, and in extreme cases to death.

1.4 Poultry production in Africa

Poultry has been domesticated for thousands of years, and has been present in Africa for centuries. Archaeological and ethnographic evidence suggests that chickens were introduced in Africa in numerous waves coming from the Mediterranean, the Red Sea and the Indian Ocean, and were then dispersed via alternative routes in the Sahara, the Horn of Africa and central and West Africa (Lyimo et al., 2014). Keeping poultry for meat and egg production has played an important socio-economic role in developing countries (Alders and Pym, 2009; Mtileni et al., 2009; Moges et al., 2010). In these countries, village chickens are not only seen as a source of animal protein and income, but they are also part of various religious and cultural traditions (Swatson et al., 2001; McAinsh et al., 2004; Muchadeyi et al., 2004; Muchadeyi et al., 2007; Alders and Pym, 2009; Mtileni et al., 2009; Moges et al., 2010).

Currently, more than 80% of the chickens raised in Africa are indigenous, and over 60% of the African rural and peri-urban families keep a flock of free range chickens in a scavenging system (Guèye, 2000; Moges et al., 2010; Pym, 2010a; Mpenda et al., 2018). Even though most poultry production in developing countries is largely focused on scavenging systems, there is another production system, high input, that focuses on using commercial breeds. The proportion of chickens belonging to these two groups can vary among countries, however in low-income countries there are more chickens from indigenous breeds than commercial breeds (Pym et al., 2006; Pym, 2010b). The distinction between these two production systems mainly lies in their management. Generally, introduced commercial breeds cannot perform at their full potential under scavenging systems, therefore they are kept under better confinement and feeding conditions. Smallholders keep their flocks of indigenous poultry in limited housing and precarious scavenging feeding conditions (Pym, 2010a, b).

Between traditional and commercial management systems, feeding resources are an important distinction. The traditional systems rely on scavenging systems, where poultry forages for water, house wastes, insects, snails, and earthworms; resources that are highly dependent on the local environment and season. On the other hand, commercial flocks are fed with on-farm mixing of complete rations of available feed ingredients, dilutions of commercial feeds, and concentrated mixtures of local ingredients with whole grains (Pym, 2010a; Ravindran, 2010). These different management and environmental conditions can expose breeds to different selective pressures (Siegel and Dunnington, 1997; Rauw et al., 1998). Studies under controlled experimental conditions have reported that temperature, humidity, and light exposure are some environmental factors that have an effect on chicken growth (Barott and Pringle, 1947; Adams et al., 1962; Smith et al., 1962; Manser, 1996; Cooper and Washburn, 1998). Generally, studies that investigate the challenges of village production systems, focus on providing detailed evaluations of the poultry performance within country, and present possible solutions and strategies that might lead to alleviating food security, such as designing breeding programs to improve local breeds (Dessie et al., 2000; Mtileni et al., 2009; Moges et al., 2010; Wondmeneh et al., 2014).

As the challenge is to increase food production, different strategies have been employed to improve chicken performance. Some strategies implemented in the past have been to control Newcastle disease by the use of a thermo-stable vaccine (Alders and Spradbrow, 2001), weaning chicks earlier to reduce brooding period (Dessie et al., 2003), and the exploitation of the genetic diversity of the local breeds (Wondmeneh, 2015). Other strategies to improve chicken production in the tropics have used crossbreeding schemes (Magothe et al., 2012), or the introduction of commercial breeds (Ngeno, 2015). A successful example of the latter can be seen in Uganda. Here, after the introduction of a commercial chicken breed into scavenging conditions of rural smallholder households, the introduced chickens grew faster and to higher weights than the indigenous chickens (Sharma et al., 2015).

Some other attempts to introduce commercial breeds have been favorable is some areas (e.g. highlands), but in warmer climates and in peri-urban production systems most attempts have failed due to the non-adaptability of the breeds, and the lack of sustainable strategies for a long-term breeding program (Philipsson et al., 2011). The problems in these cases were two-fold: first, little to no consideration was given to the environmental conditions that may affect breed adaptation, and second, there has been a lack of continuity and strategies to develop sustainable breeding programs.

African Chicken Genetic Gains (ACGG)

The African Chicken Genetic Gains (ACGG) is a project that aims to enhance the livelihoods of farmers in different African countries. On this continent, in the majority of countries, smallholder farmers produce meat and eggs for home consumption or for local markets as a source of income. The project aimed at supporting poverty reduction, enhancing food security, and empowering women farmers, by increasing the chicken productivity in the rural households. The targeted countries were Ethiopia, Nigeria and Tanzania (Figure 1.1). The ACGG initiative made available different commercial breeds of chicken to farmers taking into account their preferences. In addition to the preferences of the farmers, breeds were also chosen on their potential to withstand low-input systems in tropical conditions. Introduced chickens were tested on farm in different agro-ecologies, where the farmers kept the commercial breeds under the same management and free-range conditions as their own indigenous chickens. The aim was to build a platform for tropically adapted and farmer preferred chickens, and set up a long-term sustainable and productive chicken genetic gains program in each country.

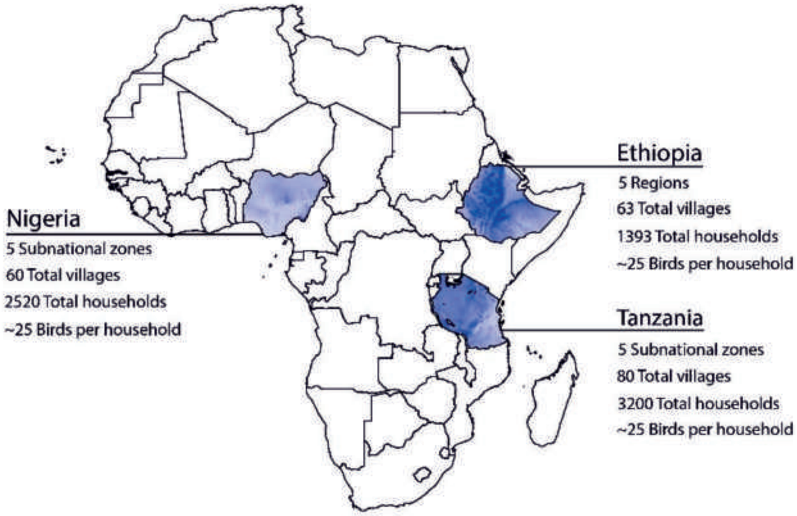

Figure 1.1 ACGG sampling scheme in Ethiopia, Nigeria and Tanzania. Darker colors indicate higher elevations.

1.5 Aim and outline of the thesis

The aim of this thesis was to understand how the environment plays a role in shaping differences in breed performance across different agro-ecological zones, and what this means for genetic improvement of chickens. To achieve this aim three objectives were targeted (Figure 1.2). The **first objective** was to use a classical breeders approach to address the interaction of genetics with the environment by estimating the genetic correlation in two different environments. The **second objective** was to develop a methodology that predicts breed suitability for specific environmental conditions. A **third objective** was to develop a methodology to predict breed phenotypic performance as a response to environmental factors.

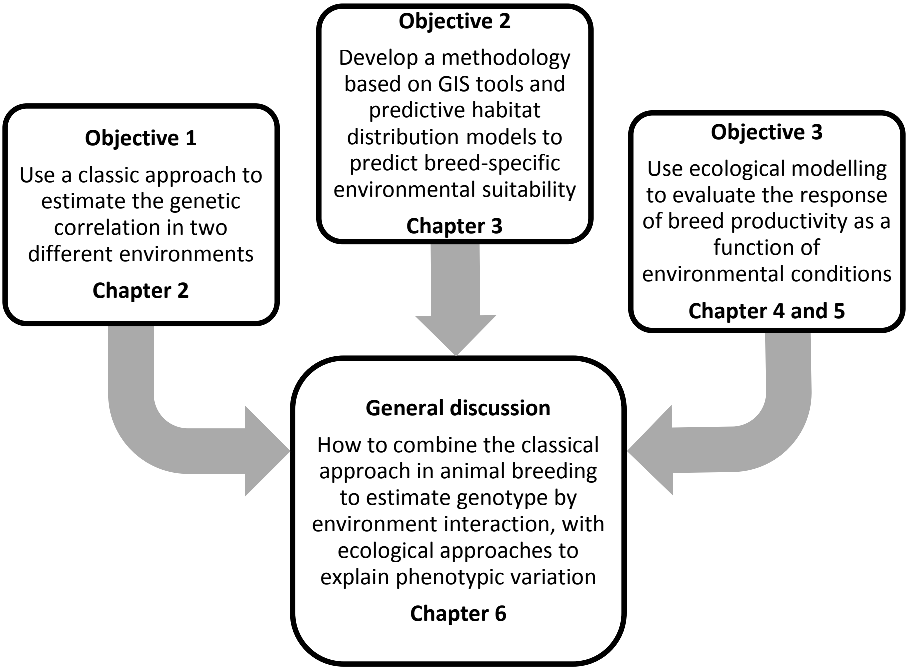

Figure 1.2 Thesis outline

To estimate the genetic correlation between traits in different environments, experimental designs need to acknowledge the mating design, and possible presence of common environmental effects. Therefore, in **chapter 2**, the optimal mating design for the estimation of the genetic correlation between environments in the presence of common environmental effects was investigated. This was approached by stochastic simulation.

Environments are dynamic, and this is not acknowledged in animal breeding approaches to explain differences in breed performance. In ecology, it is common to build models that explicitly take into account the environmental variability.

Therefore, in **chapter 3**, the aim was to develop a methodology using Geographic Information Systems (GIS) tools and predictive habitat distribution models to predict the potentially suitable range in different agro-ecological zones for different chicken breeds. In **chapter 4**, the response of productivity traits was evaluated as a function of the environment for different chicken breeds. The aim was to use phenotypic distribution models, developed for application in ecology, to establish the association between performance data and different environmental parameters. In **chapter 5**, the phenotypic distribution models from in **chapter 4** were used to evaluate if the performance of chicken breeds across different environments in one country can be used to predict the performance of the same breeds in different countries under different environmental conditions.

In the general discussion in **chapter 6**, I present the main contributions of this thesis in understanding how breeds respond differently to different environmental conditions. I make the connection between the classical approach in animal breeding of estimating genotype by environment interaction, and the use of ecological modelling to understand how the environment plays a role in shaping phenotypic variation. Integrating both approaches can lead to the development of tailor-made breeding programs that connect the genetic response in different environments to environmental variation. I discuss how the knowledge from this thesis can be implemented to improve decisions about where to introduce a breed, and to help farmers decide on which breed to choose for their own environment. I discuss why ecological modelling should be extended to different research areas in livestock.

References

Adams, R., F. Andrews, E. Gardiner, W. Fontaine, and C. Carrick. 1962. The effects of environmental temperature on the growth and nutritional requirements of the chick. Poultry Science 41: 588-594.

Alders, R., and R. Pym. 2009. Village poultry: still important to millions, eight thousand years after domestication. World's Poultry Science Journal 65: 181-190.

Alders, R., and P. Spradbrow. 2001. Controlling Newcastle disease in village chickens: a field manual. Australian Centre for International Agricultural Research (ACIAR).

Anderson, R. P., A. T. Peterson, and M. Gómez-Laverde. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98: 3-16.

Barott, H., and E. Pringle. 1947. Effect of Environment on Growth and Feed and Water Consumption of Chickens I. The Effect of Temperature of Environment During the First Nine Days After Hatch. The Journal of Nutrition 34: 53-67.

Cooper, M., and K. Washburn. 1998. The relationships of body temperature to weight gain, feed consumption, and feed utilization in broilers under heat stress. Poultry Science 77: 237-242.

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

Dessie, T., M. Tadesse, A. Yami, and K. J. Peters. 2003. Village chicken production systems in Ethiopia: 1. Flock characteristics and performance.

Elith, J. et al. 2006. Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. Ecography 29: 129-151.

Falconer, D. S. 1952. The problem of environment and selection. The American Naturalist 86: 293-298.

Guèye, E. 2000. The role of family poultry in poverty alleviation, food security and the promotion of gender equality in rural Africa. Outlook on agriculture 29: 129-136.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Joost, S. et al. 2010. Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. Animal Genetics 41: 47-63.

Joshi, J. et al. 2001. Local adaptation enhances performance of common plant species. Ecology Letters 4: 536-544.

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.

Lyimo, C. et al. 2014. Global diversity and genetic contributions of chicken populations from African, Asian and European regions. Animal genetics 45: 836-848.

Magothe, T., T. Okeno, W. Muhuyi, and A. Kahi. 2012. Indigenous chicken production in Kenya: I. Current status. World's Poultry Science Journal 68: 119-132.

Maloney, K. O., M. Schmid, and D. E. Weller. 2012. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. Methods in Ecology and Evolution 3: 116-128.

Manser, C. E. 1996. Effects of Lighting on the Welfare of Domestic Poultry: A Review. Animal Welfare 5: 341-360.

McAinsh, C. V., J. Kusina, J. Madsen, and O. Nyoni. 2004. Traditional chicken production in Zimbabwe. World's Poultry Science Journal 60: 233-246.

Michel, M. J., H. Chien, C. E. Beachum, M. G. Bennett, and J. H. Knouft. 2017. Climate change, hydrology, and fish morphology: predictions using phenotype-environment associations. Climatic change 140: 563-576.

Moges, F., A. Mellesse, and T. Dessie. 2010. Assessment of village chicken production system and evaluation of the productive and reproductive performance of local chicken ecotype in Bure district, North West Ethiopia.

Mpenda, F., M. Schilling, Z. Campbell, E. Mngumi, and J. Buza. 2018. The genetic diversity of local african chickens: A potential for selection of chickens resistant to viral infections. The Journal of Applied Poultry Research.

Mtileni, B. et al. 2009. Characterisation of production systems for indigenous chicken genetic resources of South Africa. Appl. Anim. Husb. Rural Dev 2: 18-22.

Muchadeyi, F., S. Sibanda, N. Kusina, J. Kusina, and S. Makuza. 2004. The village chicken production system in Rushinga District of Zimbabwe. Livestock Research for Rural Development 16: Retrieved June 15, 115, from http://www.lrrd.org/lrrd116/116/much16040.htm.

Muchadeyi, F. et al. 2007. Variation in village chicken production systems among agro-ecological zones of Zimbabwe. Tropical animal health and production 39: 453-461.

Mulder, H., and P. Bijma. 2005. Effects of genotype× environment interaction on genetic gain in breeding programs. Journal of Animal Science 83: 49-61.

Ngeno, K. 2015. Breeding program for indigenous chicken in Kenya (PhD Thesis). Wageningen University, Wageningen, The Netherlands.

Philipsson, J., J. Rege, E. Zonabend König, and A. Okeyo Mwai. 2011. Sustainable breeding programmes for tropical low-and medium input farming systems. In: J.M. Ojango, B. Malmfors, A.M. Okeyo (eds), Animal Genetics Training Resource, version 3, 2011. International Livestock Research Institute, Nairobi, Kenya, and Swedish University of Agricultural Sciences, Uppsala, Sweden.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological modelling 190: 231-259.

Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161-175.

Pym, R. 2010a. Poultry feed availability and nutrition in developing countries Food and Agriculture Organization of the United Nations, Poultry Development Review.

Pym, R. 2010b. Poultry genetics and breeding in developing countries Food and Agriculture Organization of the United Nations, Poultry Development Review.

Pym, R., E. Guerne Bleich, and I. Hoffmann. 2006. The relative contribution of indigenous chicken breeds to poultry meat and egg production and consumption in the developing countries of Africa and Asia. In: Proceedings of the XII European Poultry Conference. p 10-14.

Rauw, W., E. Kanis, E. Noordhuizen-Stassen, and F. Grommers. 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livestock Production Science 56: 15-33.

Ravindran, V. 2010. Poultry feed availability and nutrition in developing countries Food and Agriculture Organization of the United Nations, Poultry Development Review.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and ecology of species range limits. Annu. Rev. Ecol. Evol. Syst. 40: 415-436.

Sharma, J. et al. 2015. Higher weight gain by Kuroiler chickens than indigenous chickens raised under scavenging conditions by rural households in Uganda. Livestock Research for Rural Development 27.

Siegel, P., and E. Dunnington. 1997. Genetic Selection Strategies-Population Genetics. Poultry science 76: 1062-1065.

Smith, A. B., J. Alsdurf, M. Knapp, S. G. Baer, and L. C. Johnson. 2017. Phenotypic distribution models corroborate species distribution models: A shift in the role and prevalence of a dominant prairie grass in response to climate change. Global Change Biology 23: 4365-4375.

Smith, R., J. Kan, and T. Rokeby. 1962. The Effects of Poultry Housing on Broiler Weight, Feed Conversion, and Mortality. Poultry Science 41: 594-608.

Swatson, H., I. Nsahlai, and B. Byebwa. 2001. The status of smallholder poultry production in the Alfred district of Kwazulu-Natal, South Africa: priorities for intervention. In: Livestock Community and Environment. Proceedings of the 10th Conference of the Association of Institutions for Tropical Veterinary Medicine

Wondmeneh, E. 2015. Genetic Improvement in Indigenous Chicken of Ethiopia (PhD Thesis) Wageningen Univeristy, Wageningen, The Netherlands.

Wondmeneh, E., E. van Der Waaij, D. Tadelle, H. Udo, and J. Van Arendonk. 2014. Adoption of exotic chicken breeds by rural poultry keepers in Ethiopia. Acta Agriculturae Scandinavica, Section A—Animal Science 64: 210-216.

Chapter 2

Optimizing design to estimate genetic correlations between environments with common environmental effects

Maria Lozano-Jaramillo[1], Hans Komen[1], Yvonne CJ Wientjes[1], Han A Mulder[1], John WM Bastiaansen[1]

[1]Wageningen University & Research Animal Breeding and Genomics, P. O. Box 338, 6700 AH Wageningen, The Netherlands

Submitted

Abstract

When genotypes respond differently to changes in the environment this is defined as genotype by environment interaction (G x E). Breeding programs for different species aim to improve performance by testing members of full- and half-sib families in different environments. The presence of common environmental effects within families generates covariance between siblings, and these effects should be taken into account when estimating a genetic correlation. Therefore an optimal design should be established to accurately estimate the genetic correlation between environments. We used stochastic simulation to find the optimal population structure to accurately estimate genetic correlations between environments using a combination of full-sib and half-sib groups with different levels of common environmental effects. Results showed that ignoring common environmental effects when they are present in the population will lead to an upwards bias in the estimated genetic correlation of on average 0.3. When no common environmental effects are present in the population and the mating ratio is one female per male, the lowest standard error (SE) of the estimated genetic correlation was observed with 10 offspring per sire per environment. When common environmental effects are present in the population and are included in the model, the lowest SE is obtained with a higher number of females per sire (minimum of 5), and with a minimum number of offspring per environment of 10. Studies that aim to estimate the magnitude of G x E should acknowledge the presence of common environmental effects, and choose the mating ratio accordingly.

2.1 Introduction

The purpose of animal breeding programs is to select and breed animals that will produce more efficiently under future production settings. How well these animals perform depends on both their genotype and the production environments where they are selected and later kept (Falconer, 1952; Falconer and Mackay, 1996b). Ranking of animals may differ between environments as a result of genotype by environment interaction (G x E) (Falconer and Mackay, 1996a). The presence of G x E can be determined by estimating the genetic correlation (r_g) between the environments, using measurements of related individuals for a given trait in two environments (Falconer, 1952). Having unbiased and accurate estimates of the genetic correlation is important when predicting the response in another environment than the one where the selection took place, and when optimizing genetic improvement programs in terms of collecting phenotypic and genotypic information in production environments (Mulder and Bijma, 2005; Mulder, 2016).

Genetic correlations can be estimated based on records of relatives in two environments. Similarities between related individuals can be due to heritable genetic effects, but also due to common environmental effects. Estimating the correlation can be done from data collected in breeding programs (Falconer, 1952), from specifically designed experiments, or a combination of both (Sae-Lim et al., 2016). Estimating genetic correlations between environments based on full and half sibs in different environments is common in pigs, poultry, and fish, while in cattle estimation of genetic correlations between environments is almost entirely based on half-sibs. Full-sib or half-sib animals are initially kept together as e.g. litter groups (pigs), hatched chicks in pens (chickens), or as large groups of fry in tanks or cages (fish) until the age of weaning or individual tagging, which leads to common environmental effects. To get an accurate and unbiased estimate of the genetic correlation, this common environmental effect should be taken into account.

Predicting the standard error (SE) of the genetic correlation has been studied for many years. Robertson (1959) presented a theoretical basis for predicting the standard error (SE) of the genetic correlation for specific types of relatives (i.e. either only full-sibs (FS) or only half-sibs (HS)) in the absence of common environmental effects. Sae-Lim et al (2010) studied the standard error of the estimated genetic correlation for a specific combination of a full-sib half-sib design in which one male was mated to two different females and common environmental effects were not accounted for. Omitting the common environmental effects (c^2) can lead to severely biased estimates of the genetic parameters (Clément et al., 2001), and therefore accounting for these effects is of importance when estimating genetic correlations. Bijma and Bastiaansen (2014) presented a formula for the standard error of genetic correlation estimates taking common environmental effects into account, but their work was developed for a purebred-crossbred scenario, thereby limiting the family design to only half-sib groups.

Knowledge on the optimal full-sib-half-sib structure to minimize bias and standard error of the estimates of the genetic correlation between performance in different environments is currently lacking for situations where common environmental effects are present. The main objective of this study was therefore, to identify optimal mating designs for the estimation of genetic correlation between environments in the presence of common environmental effects. To achieve this, we used stochastic simulations to compare scenarios where the ratio of FS and HS relationships, as well as the mating ratio was varied.

2.2 Materials and methods

Experimental populations were created by stochastic simulations in R software version 3.2.2 (R Development Core Team, 2016) running in RStudio version 1.0.153 (RStudio Team, 2015). The genetic correlation between environments was estimated using an animal model implemented in ASReml (Gilmour et al., 2014). Simulation was performed with and without the presence of common environmental effects for each full-sib family, and estimation was performed with and without accounting for common environmental effects, resulting in a 2x2 design of presence or absence of common environmental effects.

2.2.1 Populations

The testing structure was based on a split-family design, where the generated FS offspring were divided equally over two environments, had trait records, and only their parents contributed to the one generation pedigree. Sae-Lim et al. (2010) suggested that to obtain an unbiased estimate of the genetic correlation, a fixed population size should be equal to 2000 individuals per environment. To compare scenarios with equal requirements for phenotyping efforts, all our designs had a constant population size of 2000 individuals per environment. The trait heritability (h^2) was 0.3 in each environment, and the genetic correlation (r_g) between environments was set equal to 0.5. The investigated variables were the number of sires (20 to 1000), number of dams per sire (1,5, or 10), and the resulting number of offspring per dam was adjusted to keep the total population size constant. All scenarios were simulated with and without common environmental effects and all datasets were analyzed with a model accounting for or ignoring the common environmental effects. The value c^2 is the variance of the common environmental effect (σ_{C}^2) as a proportion of the total phenotypic variance. The simulated values for c^2 were 0, 0.05, and 0.1 (Table 2.1).

2.2.2 Breeding values and phenotypes

True breeding values of parents for two environments were drawn from a bivariate normal distribution with means of zero, variances equal to the additive genetic variance (N(0,σ_{A}^2)), and a genetic correlation of 0.5 between the two environments. The phenotypic variance (σ_{P}^2) was set to 1. The correlation between common environmental effects for the two environments was set to 1. For each full-sib group, the common environmental effect was equal for all individuals. The common environmental effects for each FS group were simulated by drawing values from a standard normal distribution with mean of zero, and the common environmental effect variance (N(0, σ_{C}^2)).

True breeding values for each individual offspring for each environment were simulated as:

$$TBV_0 = 0.5\left(TBV_m + TBV_f\right) + ms,$$

where $\mathrm{TBV_m}$ and $\mathrm{TBV_f}$ are the true breeding values previously assigned to the male and female parents respectively, and ms is the Mendelian sampling term. The Mendelian sampling term was drawn from a normal distribution with mean of zero, and a variance equal to half the additive genetic variance (N(0, $0.5\sigma_{\mathrm{A}}^2$)). Offspring phenotypes were obtained by adding their true breeding value, common environmental effect, and a simulated environmental effect sampled from a normal distribution N(0, $(1 - \sigma_A^2 - \sigma_{\mathrm{C}}^2)$). Simulations experiments were replicated 500 times.

Table 2.1. Simulated population structures for three different mating ratios, each one with three different levels of common environmental effects (c^2).

Mating ratio	c^2	**Population structure**								
1:2	0;	Number of sires	1000	500	200	100	50	40	25	20
	0.05;	Half-sib number per environment*	2	4	10	20	40	50	80	100
	0.1	Number of dams	2000	1000	400	200	100	80	50	40
		Full-sib number per environment	1	2	5	10	20	25	40	50
1:5	0;	Number of sires	400	200	100	80	40	20		
	0.05;	Half-sib number per environment*	5	10	20	25	50	100		
	0.1	Number of dams	2000	1000	500	400	200	100		
		Full-sib number per environment	1	2	4	5	10	20		
1:10	0;	Number of sires	200	100	50	25	20			
	0.05;	Half-sib number per environment*	10	20	40	80	100			
	0.1	Number of dams	2000	1000	500	250	200			
		Full-sib number per environment	1	2	4	8	10			

*Sire offspring per environment

2.2.3 Estimation of genetic parameters

The genetic correlation was estimated from the simulated data using two different models, one accounting for the presence of common environmental effects and another ignoring these effects. The models were fitted as follows:

Model 1

$$\begin{bmatrix}\mathbf{y_1}\\ \mathbf{y_2}\end{bmatrix} = \begin{bmatrix}\mathbf{x_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{x_2}\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z_2}\end{bmatrix}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix} + \begin{bmatrix}\mathbf{W_1}\\ \mathbf{W_2}\end{bmatrix}\mathbf{c} + \begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\end{bmatrix} \tag{2.1}$$

Model 2

$$\begin{bmatrix}\mathbf{y_1}\\ \mathbf{y_2}\end{bmatrix} = \begin{bmatrix}\mathbf{x_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{x_2}\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z_2}\end{bmatrix}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix} + \begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\end{bmatrix} \tag{2.2}$$

where $\mathbf{y_1}$ and $\mathbf{y_2}$ are vectors with the phenotypes measured in environment 1 and 2, $\mathbf{x_1}$ and $\mathbf{x_2}$ are the incidence vectors relating the traits to the mean in environment 1 (μ_1) or environment 2 (μ_2), $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices relating the phenotypes to the random additive genetic effect in environments 1 ($\mathbf{a_1}$) and 2 ($\mathbf{a_2}$), $\mathbf{W_1}$ and $\mathbf{W_2}$ are the incidence matrices relating the phenotypes per sire offspring to the common environmental effect ($\mathbf{c}$), and $\mathbf{e_1}$ and $\mathbf{e_2}$ are the vectors containing the random residual effects. The residual covariance was fixed at zero, because each animal is performing in only one environment.

2.2.4 Summarizing simulation output

The estimates and standard errors were obtained as reported by ASReml version 4.1 software (Gilmour et al., 2014) for each of the genetic parameters h_1^2, h_2^2, r_g, c^2. Replicates with estimates of r_g in the -1 to 1 range, that had converged, and where the variance-covariance matrices were positive definite were kept for further analysis (Table S2.1).

2.2.5 Accuracy of genetic correlation estimates

The accuracy of the estimated genetic correlation was obtained by 1) the standard deviation across estimates from different simulation replicates and 2) the average of the standard errors reported by ASReml. For some scenarios, the standard errors of the genetic correlation were compared to deterministic predictions. First, for scenarios with only full-sibs, or only half sibs the SE was predicted by Robertson's equation (1959):

$$SE(\widehat{r_g}) \approx \sqrt{\frac{\left[1 + nt\left(1 - r_g^2\right)\right]^2 + r_g^2}{(N-1)n^2t^2}} \tag{2.3}$$

Where N is the number of sire families, n is the number of offspring per environment and t is the intra-class correlation (e.g. for half-sibs, t = $0.25h^2$), and r_g is the known genetic correlation between environments. Second, for scenarios with only half-sibs the SE was predicted using the equation from Bijma and Bastiaansen (2014):

$$SE(\widehat{r_g}) \approx \sqrt{\frac{\frac{1}{{\rho_x}^2{\rho_y}^2} + (1 + \frac{0.5}{{\rho_x}^4} + \frac{0.5}{{\rho_y}^4} - \frac{2}{{\rho_x}^2} - \frac{2}{{\rho_y}^2})r_g^2 + r_g^4}{N-1}} \tag{2.4}$$

Where N is the number of sire families, ${\rho_x}^2$ and ${\rho_y}^2$ are the reliabilities of estimated breeding values in each environment, and r_g is the true genetic correlation between environments.

2.3 Results

2.3.1 Standard error of estimated r_g with c^2= 0

The average standard errors as reported by ASReml were not different from the standard deviation estimated over replicates. Therefore, we only reported the latter. The SE of the $\widehat{r_g}$ for the scenarios where equations from Robertson (1959) and Bijma and Bastiaansen (2014) could be applied (i.e. only FS or only HS), did not differ from SE obtained by the simulations (see Table S2.2 for the SEs). Only results from simulations are reported here. In the absence of common environmental effects, the smallest standard error was obtained with full-sib groups only, where the minimum standard error for our population size of 2000 per environment (0.097) was found for 10 offspring per sire per environment (Figure 2.1). With half-sib groups only, the minimum standard error (0.144) was obtained with 20 offspring per sire per environment (Figure 2.1).

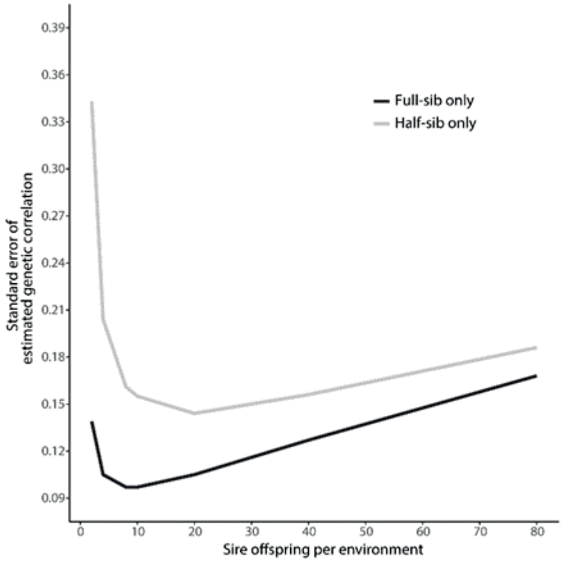

Figure 2.1 Standard error of the estimated genetic correlation for simulated scenarios for only half-sibs and only full-sibs for a population size of 2000 individuals per environment and no common environmental effects (c^2).

In the absence of common environmental effects, having a FS-HS structure gave higher SE of the $\widehat{r_g}$ than having only FS, but lower SE than having only HS (Figure 2.1 and Figure 2.2a). With 10 offspring per sire per environment, the SE of the $\widehat{r_g}$ increased from 0.097 in a 1:1 mating ratio (Figure 2.1) to 0.100, 0.114, and 0.128 for a 1:2, 1:5, and 1:10 ratio respectively (Figure 2.2a). When the number of offspring per sire per environment increased, and consequently the number of sires decreased, the SE increases rapidly with a 1:1 mating ratio (Figure 2.1) but much slower with more than one female per male (Figure 2.2a). When the number of offspring per sire per environment was equal or greater than 40, the SE for a full-sib family design was larger than the SE for all the other mating ratios (Figure 2.2a). In summary, in the absence of common environmental effects, a 1:1 mating ratio with 10 offspring per sire per environment will result in the smallest SE.

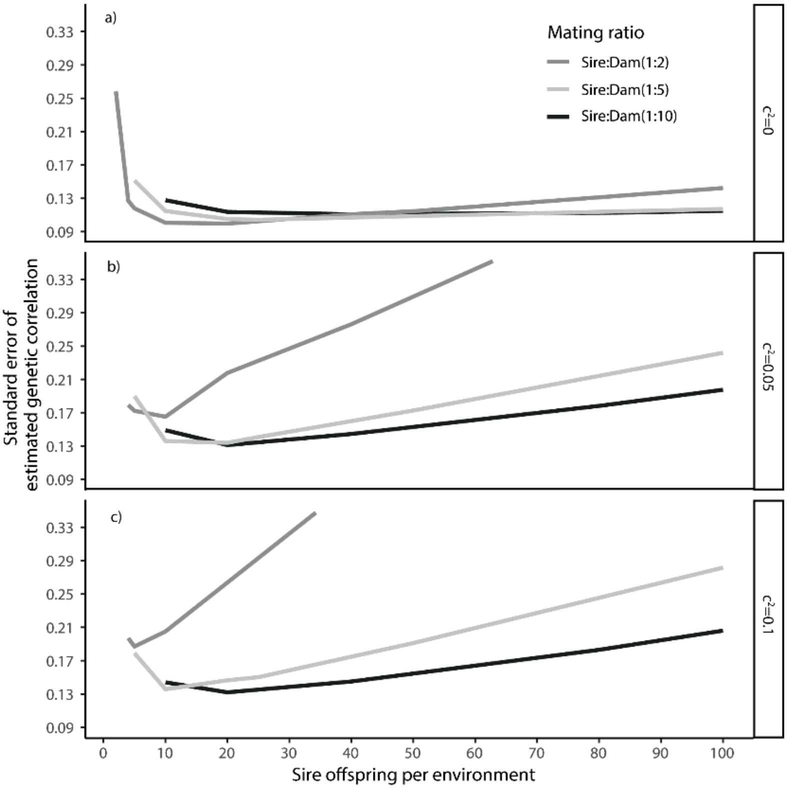

Figure 2.2 Standard error of the estimated genetic correlation for different simulated scenarios using the correct model for three mating ratios for a population size of 2000 individuals per environment, and three levels of common environmental effects (c^2). In panel a) c^2=0 and no common environmental effects included in the model, in panel b) c^2=0.05 and the common environmental effects are included in the model, and in panel c) c^2=0.1 and the common environmental effects are included in the model.

2.3.2 Standard error of estimated r_g with c^2 > 0

In the presence of c^2 in the population, and when accounting for c^2 in the model, the smallest SE of the $\hat{r}_g$ for all mating ratios was found when the number of offspring per sire per environment was equal to 20 (Figure 2.2b and Figure 2.2c). The lowest SE of the $\hat{r}_g$ was reached when the mating ratio was 1:10 with 20 offspring per sire per environment, which resulted in each offspring having 1 full-sib and 18 half-sibs in the same environment (SE=0.131; Figure 2.2b and 2.2c; Table 2.1). When c^2 was present, and included in the model, the SEs of the $\hat{r}_g$ increased for all scenarios, compared to the scenarios when c^2 was not present and not included in the model (Figure 2.2). When the number of offspring per sire per environment increased, the standard error for the 1:2 mating ratio became extremely large. However, when there were more females mated per male, the standard error of the $\widehat{r_g}$ also increased when the number of offspring per sire per environment increased, but at a much lower rate (Figure 2.2). In the presence of c^2, 10 females mated per male with 20 offspring per sire per environment resulted in the smallest SE.

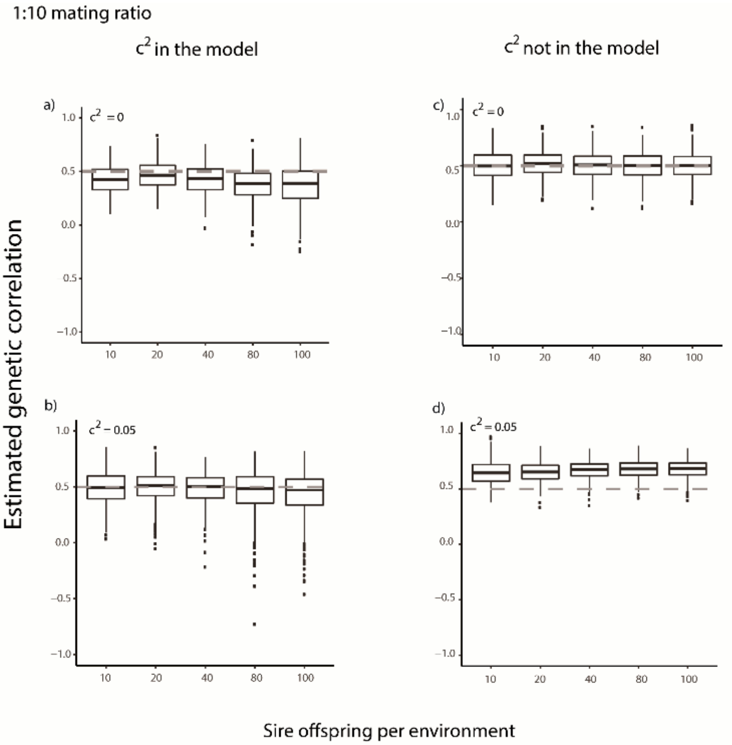

Figure 2.3 Distribution of the estimated genetic correlation for 1:10 mating ratio for a population size of 2000 individuals per environment when a) c^2=0 and b) c^2=0.05 and common environmental effects are included in the model. When c) c^2=0 and d) c^2=0.05 and common environmental effects are not included in the model. The dashed line indicates the simulated value of the genetic correlation (r_g=0.5).

2.3.3 Estimates of r_g and c^2 from the correct model

In the absence of c^2 and without the common environmental effect in the model, the $\widehat{r_g}$ were unbiased (Figure 2.3c). When c^2> 0, and the common environmental effect was included in the model, the $\widehat{r_g}$ were also unbiased (Figure 2.3b). This pattern was the same for all different mating ratios and levels of c^2 tested. We only show the results for the 1:10 mating ratio. The estimates of c^2 themselves were also unbiased when common environmental effects were present in the population and included in the model (Figure 2.4b).

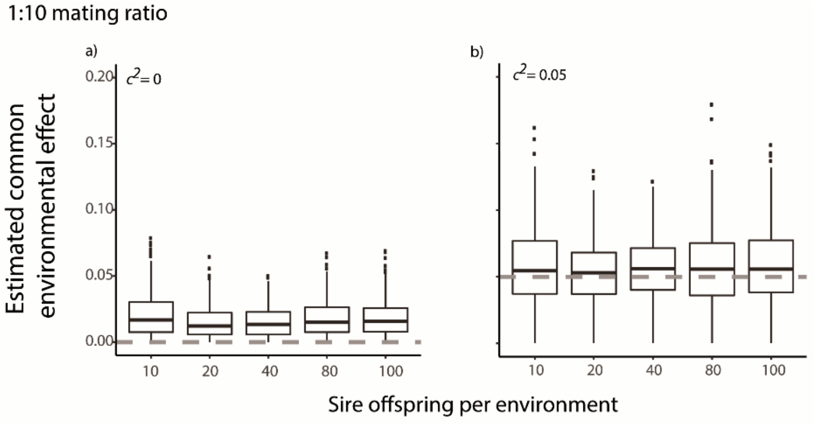

Figure 2.4 Distribution of the estimated common environmental effect for 1:10 mating ratio for a population size of 2000 individuals per environment when a) c^2=0 and when b) c^2=0.05. The dashed line indicates the simulated value of the common environmental effect c^2=0.0, and c^2=0.05 respectively.

2.3.4 Estimates of r_g and c^2 from the incorrect model

In the absence of c^2, and with the common environmental effects present in the model there was a downward bias of the $\hat{r_g}$ of on average 0.1 (Figure 2.3a). On the other hand, estimates of genetic correlations showed a upward bias of on average 0.3 when common environmental effects were present in the population, but not included in the model (Figure 2.3d). This pattern was the same for all different mating ratios and levels of c^2 tested. We only show the results for the 1:10 mating ratio. The bias in $\hat{r_g}$ increased when c^2 was not included in the model, while present in the population. The estimates of c^2were biased upward by on average 0.02 when the common environmental effects were not present in the population and included in the model (Figure 2.4a).

2.4 Discussion

The aim of this study was to identify the optimal mating design to estimate the genetic correlation between environments in the presence of common environmental effects and different full-sib and half-sib ratios. Estimates of genetic correlations of traits between different environments are of crucial importance for the design and optimization of breeding programs. Neglecting the existence of a G x E interaction will lead to reduced response to selection (Mulder and Bijma, 2005; Dominik and Kinghorn, 2008). For accurate estimates, experimental designs should be optimized. In many situations the populations used to collect data can be designed to a certain degree.

Some studies can, or have to use field data to estimate genetic correlations (Mulder et al., 2004; Haile-Mariam et al., 2015; Sevillano et al., 2016; Godinho et al., 2018), but other studies use experimental designs under specific population structures to estimate the magnitude of G x E (Lwelamira, 2012; Dottavio et al., 2019; Lillehammer et al., 2019). As common environmental effects generate similarities between individuals, these effects should be taken into account when designing the best population structure to estimate the genetic correlations (Winkelman and Peterson, 1994). Here we show the importance of choosing the experimental design and taking common environmental effects into account. Not accounting for these effects will lead to biased estimates of the genetic correlation and common environmental effects will increase the standard error of genetic correlations.

2.4.1 Levels of c^2

Siblings often share a common environment during the early rearing stage, increasing the phenotypic covariance between relatives. In livestock, some studies have reported high levels of c^2 ranging from 0.06 to 0.5. For example in several pig breeds, the proportion of variance due to common environmental effects has been reported to be around 0.34 for intramuscular fat, between 0.06-0.53 for fat composition in Iberian pig lines (Ibáñez-Escriche et al., 2016), and between 0.22-0.26 for days to 100kg (Li and Kennedy, 1994). For catfish body weight, estimates reported range between 0.06-0.33 (Tran, 2016). For Venda chicken body weight, common environmental effects for weight at hatching and weight at 4 weeks were reported to be 0.39 and 0.18 respectively (Norris and Ngambi, 2006). Here we show that the higher the level of c^2, the greater the bias of the estimated genetic correlation when not including common environmental effects in the model. Therefore, the high levels of c^2 in these studies highlight the relevance to take common environmental effects into account in the model and to optimize the design for estimating genetic correlations.

2.4.2 Optimizing mating designs

When the number of animals to be phenotyped is fixed, the standard error of the estimated genetic correlation can still be influenced by the mating design. In animal breeding programs, nested designs are traditionally used, where one male is mated to a different number of females. Bijma and Bastiaansen (2014) predicted the SE of the estimated genetic correlation for mating ratios that are typically used for four species under different levels of c^2. Comparing mating ratios of 1:2 for tilapia, 1:7 for laying hens, 1:10 for pigs, and 1:12 for broilers, they report the highest standard error of the estimated genetic correlation for a 1:2 design such as used in Nile tilapia. Increasing the number of females per male was found to lower the SE of the $\widehat{r_g}$. This is in agreement with our study, where we show that having a higher number of females per male will give more accurate estimates of the genetic correlation regardless of the level of c^2. Even though having more than 10 females per male may give even smaller SE of the $\widehat{r_g}$ this will reduce the number of males used considerably, which is typically not desired in breeding programs for reasons of genetic diversity, when the experiment is part of a breeding program.

With a given population size of e.g. 2000 individuals per environment, having a higher number of females per male reduces the number of full-sibs (offspring per dam) per environment, but keeps the number of half-sibs (offspring per sire) per environment the same. Distinguishing common environmental effects from genetic effects requires that related individuals exist in different environments, and that there is variation in the degree of relatedness between individuals that share the particular common environment (Kruuk and Hadfield, 2007). To minimize bias in the estimates of the genetic correlation, having more dams per sire should be implemented.

In cases when the number of sire offspring per environment increases, the number of parents linking this offspring to each environment decreases. Therefore, when the number of offspring per sire is larger than 40, the number of females per sire should be increased to be at least 5. With common environmental effects present in the population, the optimum number of offspring per sire per environment is equal to 20, decreasing the number of sires accordingly. Having more dams per sire will then be needed to compensate for the low number of sires.

In accordance with Robertson (1959), in the absence of common environmental effects, the 1:1 mating ratio gives the lowest SEs for the estimated genetic correlation. However, to benefit from using a 1:1 mating ratio, there has to be certainty about the absence of common environmental effects, so that there is no bias on the estimated genetic correlation. Under this mating ratio, the variance between full-sibs is half of the additive genetic variance plus the variance due to common environmental effects. Ignoring the common environmental effects will therefore overestimate the additive genetic variance by two times the common environmental effect variance (Kruuk and Hadfield, 2007). On the other hand, using only half-sibs, gives the highest SEs for the estimated genetic correlation. It should be noted that the SEs of the estimated genetic correlation obtained from the Bijma and Bastiaansen (2014) formula are slightly lower for all the different population structures compared to the SEs obtained by the simulations and by Robertson (1959) formula. The formula by Bijma and Bastiaansen (2014), was derived under some assumptions, and therefore can be interpreted, as the authors suggest, as a lower bound of the SE of the $\widehat{r_g}$ for only half-sib relationships.

Nested designs, with a 1:2 mating ratio, are commonly seen in for instance Nile tilapia breeding programs. These designs are applied because full-sibs need to be reared separately until tagging before they are transferred to communal growth environment. However, family production in some fish breeding programs relies on natural mating and families are kept separate for a long period, making it likely that common environmental effects are introduced (Winkelman and Peterson, 1994; Trọng et al., 2013). Two dams per sire is the minimum design to allow estimation of c^2. However, different G x E studies have shown difficulties when trying to estimate genetic parameters while including common environmental effects in the model. They attribute this problem to the low number of dams mated to each sire (1:2 or 1:3 mating ratio), and to a fixed effect such as spawning date being confounded with the c^2 (Trọng et al., 2013; Omasaki et al., 2016). Nonetheless, other studies have estimated significant levels of common environmental effects present for different tilapia species (Thodesen et al., 2013; Thoa et al., 2015). Using a 1:2 mating design in Red Tilapia, estimates for c^2 ranging from 0.23-0.59 were found in fish kept in ponds, and estimates between 0.1-0.31 were found in fish kept in cages (Nguyen, Azhar and Thoa, 2017). These significant levels of c^2 and the fact that different studies had trouble in estimating c^2 highlight the importance of optimizing designs to estimate both c^2 and r_g.

2.4.3 Impact of biased *rg* estimates

Biased genetic parameter estimates will lead to an erroneous conclusion about the genetic values of individuals for the different environments, to over- or underestimating the importance of G x E, and possibly suboptimal designed breeding programs. Biased genetic correlation estimates can also lead to erroneous conclusions about the genetic improvement in breeding programs (Dominik and

Kinghorn, 2008; Chu et al., 2018). Overestimated genetic correlations would result from wrongly ignoring the presence of c^2, and could lead to the conclusion that G x E is absent, when in fact it exists. These biased estimates of genetic correlations may cause the decision to have a single breeding program to serve the different environments, while the optimal decision would be to have multiple separate breeding programs (Mulder et al., 2006b).

Taking common environmental effects into account is not systematically done when evaluating a genotype by environment interaction, as they can be difficult to estimate (Kolstad et al., 2006; Omasaki et al., 2016). Here we show the importance of estimating these effects, as they also have an influence on the bias and accuracy of estimating genetic parameters. In cases where there is an uncertainty about the presence of common environmental effects in the population, but the effects are included in the model while being equal to zero, there is a slight downward bias of the estimated genetic correlation. Therefore a proper experimental design should be implemented to estimate the common environmental effects accurately and the effect should be included in the model if present in the population.

Robertson (1959) presented a formula to predict the standard error of the genetic correlation for restricted family groups, either only full-sibs or only half-sibs in the absence of common environmental effects. Later on, Bijma and Bastiaansen (2014) presented a formula to estimate the standard error of the genetic correlation accounting for common environmental effects, but restricted the family design to only-half sib group. In animal breeding it is common to estimate the genetic correlation between environments for a combination of full-sibs and half-sibs separated in two environments. With these family designs the presence of common environmental effects is very likely. Therefore, the previous theoretical bases should be considered with some caveats, as they do not resemble the actual genetic improvement programs. This study, acknowledges the combination of full-sibs and half-sibs groups and the presence of common environmental effects for the estimation of the standard error of the genetic correlation. This will allow for animal breeders to decide on the appropriate population structure needed to achieve unbiased estimates under more realistic scenarios.

This simulation study focused on finding the optimal population design to minimize the standard error of the genetic correlation in the presence of common environmental effects. The constant population size of 2000 individuals per environments allowed us to make comparisons across different mating ratios. We show that mating structures with more than 2 females per male should be used to better disentangle the common environmental effects from the genetic effects, thereby obtaining a lower SE of the estimated genetic correlation. Only when there is certainty about the absence of common environmental effects in the population, a single female per male will give the lowest SE and an unbiased estimate of the genetic correlation.

2.5 Acknowledgements

The authors sincerely thank the Koepon Foundation for providing a scholarship to the first author.

4.6 Supplementary information

Table S4.1 Number of iterations that converged for all the simulated scenarios.

		Population structure															
		Half-sib number per environment*															
		2		4		10		20		40		50		80		100	
Mating ratio	c^2	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model
1:2	0	488	500	477	500	467	500	445	500	427	500	414	500	360	500	350	500
	0.05	374	500	474	500	484	500	458	500	465	500	452	500	425	500	402	500
	0.1	430	500	498	499	499	500	487	500	454	500	431	500	407	500	362	500

		Population structure											
		Half-sib number per environment*											
		5		10		20		25		50		100	
Mating ratio	c^2	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model
1:5	0	495	500	500	500	500	500	500	500	497	500	487	500
	0.05	480	500	493	500	492	500	495	500	486	500	460	500
	0.1	495	500	485	500	500	500	500	500	497	500	487	500

		Population structure									
		Half-sib number per environment*									
		10		20		40		80		100	
Mating ratio	c^2	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model	In the model	Not in the model
1:10	0	277	500	254	500	244	496	271	497	262	497
	0.05	471	500	494	495	491	497	484	497	474	500
	0.1	465	500	460	500	480	484	493	500	490	500

*Sire offspring per environment

Table S4.2 Standard errors of the estimated genetic correlation for the scenarios based on theory and simulations for only full-sibs or only half-sibs for a population size of 2000 individuals per environment and no common environmental effects (c^2).

	Sire offspring per environment						
Scenario	2	4	8	10	20	40	80
Robertson (FS)	0.14	0.114	0.104	0.103	0.11	0.131	0.17
Simulation (FS)	0.139	0.105	0.099	0.097	0.105	0.127	0.168
Robertson (HS)	0.257	0.197	0.162	0.155	0.146	0.157	0.188
Bijma & Bastiaansen (HS)	0.239	0.185	0.153	0.147	0.141	0.153	0.187
Simulation (HS)	0.343	0.204	0.161	0.155	0.144	0.156	0.186

References

Bijma, P., and J. W. Bastiaansen. 2014. Standard error of the genetic correlation: how much data do we need to estimate a purebred-crossbred genetic correlation? Genetics Selection Evolution 46: 79.

Chu, T. T., J. W. M. Bastiaansen, E. Norberg, and P. Berg. 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 68: 1-10.

Clément, V. et al. 2001. Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects. Genetics Selection Evolution 33: 369.

Dominik, S., and B. Kinghorn. 2008. Neglecting genotype × environment interaction results in biased predictions from selection index calculations. Livestock Science 114: 233-240.

Dottavio, A. et al. 2019. Interacción genotipo x estación del año sobre caracteres de producción de carne en pollos camperos. Compendio de Ciencias Veterinarias 9: 15-21.

Falconer, D. S. 1952. The problem of environment and selection. The American Naturalist 86: 293-298.

Falconer, D. S., and T. F. Mackay. 1996a. Introduction to Quantitative Genetics, Longman group.

Falconer, D. S., and T. F. C. Mackay. 1996b. Introduction to Quantitative Genetics, Longman group.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2014. ASReml User Guide Release 4.1 Structural Specification. VSN International Ltd.

Godinho, R. et al. 2018. Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. Journal of animal science 96: 817-829.

Haile-Mariam, M., J. Pryce, C. Schrooten, and B. Hayes. 2015. Including overseas performance information in genomic evaluations of Australian dairy cattle. Journal of dairy science 98: 3443-3459.

Ibáñez-Escriche, N., E. Magallón, E. Gonzalez, J. Tejeda, and J. Noguera. 2016. Genetic parameters and crossbreeding effects of fat deposition and fatty acid profiles in Iberian pig lines. Journal of Animal Science 94: 28-37.

Kolstad, K., I. Thorland, T. Refstie, and B. Gjerde. 2006. Genetic variation and genotype by location interaction in body weight, spinal deformity and sexual maturity in Atlantic cod (Gadus morhua) reared at different locations off Norway. Aquaculture 259: 66-73.

Kruuk, L., and J. Hadfield. 2007. How to separate genetic and environmental causes of similarity between relatives. Journal of Evolutionary Biology 20: 1890-1903.

Li, X., and B. Kennedy. 1994. Genetic parameters for growth rate and backfat in Canadian Yorkshire, Landrace, Duroc, and Hampshire pigs. Journal of Animal Science 72: 1450-1454.

Lillehammer, M. et al. 2019. Genetic parameters of resistance to amoebic gill disease in two Norwegian Atlantic salmon populations. Aquaculture 508: 83-89.

Lopes, M. S., J. W. M. Bastiaansen, L. Janss, E. F. Knol, and H. Bovenhuis. 2015. Estimation of Additive, Dominance, and Imprinting Genetic Variance Using Genomic Data. G3: Genes|Genomes|Genetics 5: 2629-2637.

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.

Mulder, H., and P. Bijma. 2005. Effects of genotype× environment interaction on genetic gain in breeding programs. Journal of Animal Science 83: 49-61.

Mulder, H. A. 2016. Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. Frontiers in Genetics 7.

Mulder, H. A., A. F. Groen, G. De Jong, and P. Bijma. 2004. Genotype×Environment Interaction for Yield and Somatic Cell Score with Automatic and Conventional Milking Systems. Journal of Dairy Science 87: 1487-1495.

Mulder, H. A., R. F. Veerkamp, B. J. Ducro, J. A. M. van Arendonk, and P. Bijma. 2006. Optimization of Dairy Cattle Breeding Programs for Different Environments with Genotype by Environment Interaction. Journal of Dairy Science 89: 1740-1752.

Norris, D., and J. Ngambi. 2006. Genetic parameter estimates for body weight in local Venda chickens. Tropical Animal Health and Production 38: 605-609.

Omasaki, S. K., H. Charo-Karisa, A. K. Kahi, and H. Komen. 2016. Genotype by environment interaction for harvest weight, growth rate and shape between monosex and mixed sex Nile tilapia (*Oreochromis niloticus*). Aquaculture 458: 75-81.

R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved on 8 December 2016 from http://www.R-project.org.

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

RStudio Team. 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, USA. Retrieved on 30 June 2015 from http://www.rstudio.com/.

Sae-Lim, P., H. Komen, and A. Kause. 2010. Bias and precision of estimates of genotype-by-environment interaction: A simulation study. Aquaculture 310: 66-73.

Sae-Lim, P., B. Gjerde, H. M. Nielsen, H. Mulder, and A. Kause. 2016. A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. Reviews in Aquaculture 8: 369-393.

Sevillano, C. A., H. A. Mulder, H. Rashidi, P. K. Mathur, and E. F. Knol. 2016. Genetic variation for farrowing rate in pigs in response to change in photoperiod and ambient temperature. Journal of Animal Science 94: 3185-3197.

Thoa, N. P. et al. 2015. Genetic variation in survival of tilapia (*Oreochromis niloticus*, Linnaeus, 1758) fry during the early phase of rearing in brackish water environment (5–10ppt). Aquaculture 442: 112-118.

Thodesen, J. et al. 2013. Genetic improvement of tilapias in China: Genetic parameters and selection responses in growth, pond survival and cold-water tolerance of blue tilapia (*Oreochromis aureus)* after four generations of multi-trait selection. Aquaculture 396-399: 32-42.

Tran, P. H. 2016. Analysis of crosses between three strains of striped catfish (*Pangasianodon hypophthalmus*) in Viet Nam, Norwegian University of Life Sciences, Ås.

Trọng, T. Q., H. A. Mulder, J. A. van Arendonk, and H. Komen. 2013. Heritability and genotype by environment interaction estimates for harvest weight, growth rate, and shape of Nile tilapia (*Oreochromis niloticus*) grown in river cage and VAC in Vietnam. Aquaculture 384: 119-127.

Winkelman, A. M., and R. G. Peterson. 1994. Heritabilities, dominance variation, common environmental effects and genotype by environment interactions for weight and length in chinook salmon. Aquaculture 125: 17-30.

Chapter 3

Use of geographic information system tools to predict animal breed suitability for different agro-ecological zones

Maria Lozano-Jaramillo[1], John WM Bastiaansen[1], Tadelle Dessie[2], Hans Komen[1]

[1]Wageningen University & Research Animal Breeding and Genomics, P. O. Box 338, 6700 AH Wageningen, The Netherlands; [2]International Livestock Research Institute, P. O. Box 5689 Addis Ababa, Ethiopia

animal (2018) doi.org/10.1017/S1751731118003002

Abstract

Predicting breed specific environmental suitability has been problematic in livestock production. Native breeds have low productivity but are thought to be more robust to perform under local conditions than exotic breeds. Attempts to introduce genetically improved exotic breeds are generally unsuccessful, mainly due to the antagonistic environmental conditions. Knowledge of the environmental conditions that are shaping the breed would be needed to determine its suitability to different locations. Here, we present a methodology to predict the suitability of breeds for different agro-ecological zones using Geographic Information Systems (GIS) tools and predictive habitat distribution models. This methodology was tested on the current distribution of two introduced chicken breeds in Ethiopia: the Koekoek, originally from South Africa, and the Fayoumi, originally from Egypt. Cross-validation results show this methodology to be effective in predicting breed suitability for specific environmental conditions. Furthermore, the model predicts suitable areas of the country where the breeds could be introduced. The specific climatic parameters that explained the potential distribution of each of the breeds were similar to the environment from which the breeds originated. This novel methodology finds application in livestock programs, allowing for a more informed decision when designing breeding programs and introduction programs, and increases our understanding of the role of the environment in livestock productivity.

3.1 Introduction

Indigenous breeds are exposed to natural selection processes that allow them to acquire qualities that make them better suited to their environment. Native breeds have been described to be locally adapted to specific environmental conditions, as well as tolerant to different parasites and diseases (Solti et al., 2000; Köhler-Rollefson et al., 2009; Mirkena et al., 2010). Exotic breeds show an advantage in production over the indigenous breeds as they have been selected for high productivity for many generations. For this reason, many introduction programs aim to increase local egg and meat productivity in chicken, to increase wool yield in sheep, meat quality in cattle and in goats, as well as milk yield in cows. However, most programs were not successful, mainly because of the non-adaptability of the exotic breeds to the challenging tropical environments (Kosgey et al., 2006; Mirkena et al., 2010; Wurzinger et al., 2011; Haftu Kebede Sebho, 2016). What is needed are methods to predict which areas are suitable, in terms of environmental conditions, for the introduction of different breeds. Such methodology would make introduction programs and design of breeding programs more efficient.

Predictive habitat distribution models are Geographic Information Systems (GIS)-based tools that use the current climatic conditions of a species to make predictions of the potential distribution of the species (Pearson and Dawson, 2003; Hijmans and Graham, 2006; Soberón and Nakamura, 2009). These tools explain naturally occurring plant and animal distribution patterns, or assess the impact of climate change on their distributions. These models may be useful as a new tool to predict whether the introduction of a specific livestock breed has the potential to be successful, based on local climatic conditions. In livestock research, GIS tools have been used to map suitable territories for land use (Malafant, 1998; Kalivas and Apostolopoulos, 2005), to view patterns of disease transmission (Cringoli et al., 2007), and to establish conservation priorities (Bertaglia et al., 2007). However, these tools have not been used to predict suitable areas for particular livestock breeds. Neither have they been used to understand the environmental factors that may influence changes in productivity between environments.

Here we present a methodology that uses GIS tools to develop predictive habitat distribution models that can be used to predict the suitability of a breed for a particular region based on climatic information. The methodology was tested on two introduced poultry breeds in Ethiopia. Ethiopia was considered suitable for testing the methodology because it is an ecologically diverse country with a broad range of contrasting agro-ecologies defined by altitude, temperature and rainfall (Mengistu, 2003). The methodology was used to (1) make predictions on the potential suitable habitat range for each breed, (2) indicate which bio-climatic and land cover variables explain the differences between the areas predicted to be suitable for the different breeds, and (3) establish a ranking of the suitability of the two available breeds for each region. This novel methodology finds application in livestock programs, allowing a more informed decision making for the design of breeding programs and introduction programs, and increases our understanding of the role of the environment in livestock productivity.

3.2 Materials and methods

Using distribution models and GIS tools we developed a methodology and applied it to predict areas of potential suitability for two different livestock breeds. To validate the methodology we chose two different chicken breeds that are currently kept in Ethiopia. The development of the methodology involved building distribution models based on climate for each breed. Validation was done by cross validation to determine if the model could differentiate areas where the breeds are kept from areas where the breed is not present.

3.2.1 Distribution model building

To build the distribution models, we used the maximum entropy algorithm implemented by Maxent (Phillips et al., 2006). Maxent is one of the most commonly used tools in ecology to predict species distributions. It has been shown to have greater predictive power than other tools, particularly for small data sets (Elith et al., 2006). Maxent is a machine learning algorithm that uses presence-only data to relate the environmental variables and occurrence points to establish a probability of potential geographic suitability (Phillips et al., 2006; Phillips and Dudík, 2008). The output, is the probability of suitability for all map positions that can be represented as a heat map.

3.2.2 Environmental data

Ethiopia was chosen because it is a diverse country divided in nine regional states (Figure 3.1a) and five agro-ecological zones based on rainfall and elevation, the latter being a determinant for agricultural land use due to its influence on temperature (Mengistu, 2003; Table 1; Deressa et al., 2010). Additionally, to better characterize the country’s temperature and moisture regimes, a system of zonation was developed generating 18 major zones (Deressa et al., 2010; Fig. 3.1b).

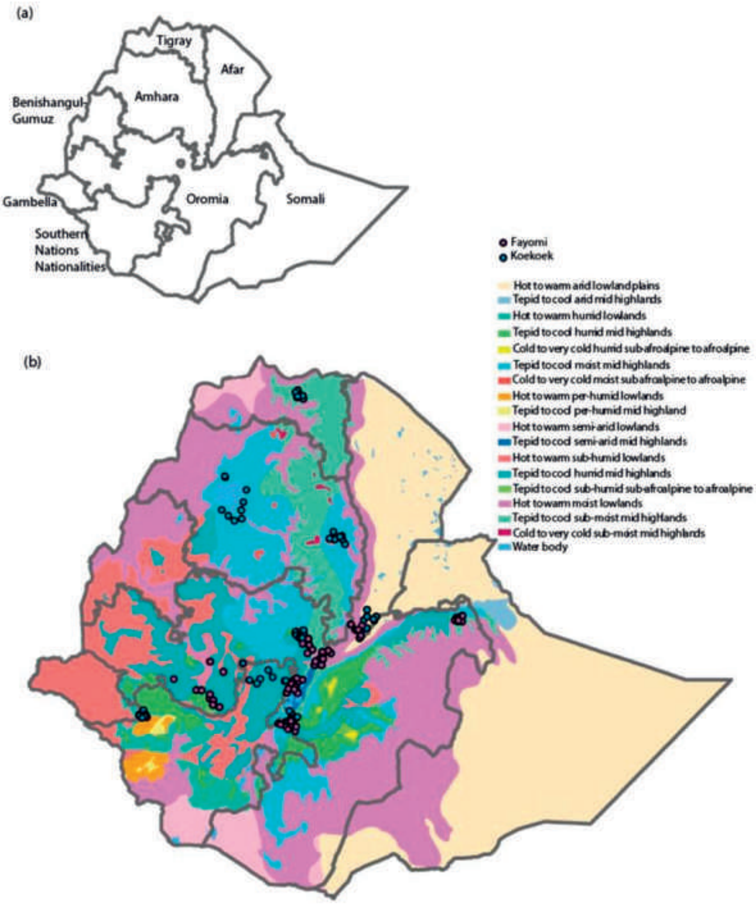

Figure 3.1 Map of Ethiopia showing (a) its nine regional states, and (b) the 18 major agro ecological zones based on temperature and precipitation. Dots on the maps indicate the localities from each of the breeds that were used to build the models.

Table 3.1 Traditional agro-ecological zones in Ethiopia.

Zone	Elevation (m)	Mean Annual precipitation (mm)	Average annual temperature (°C)
Bereha (dry-hot/desert)	< 500	< 200	> 27.5
Kola (sub-moist warm/lowlands)	500 - 1500	200 - 800	20.0 - 27.5
Weinadega (dry- warm/mid highlands)	1500-2500	800 - 1200	17.0 – 20.0
Dega (cold/highlands)	2500-3500	900-1200	11.5-17.0
Wurch (very cold or alpine/upper highlands)	> 3500	900 – 2200	< 11.5

Chickens are part of the Ethiopian village production systems, where they rely on scavenging for survival. Food source is dependent on seasonality, which in Ethiopian agricultural circumstances is strongly related to temperature and rainfall. We chose the sets of environmental variables that would represent trends in seasonality (temperature and precipitation); variables that would have an influence on, or would reflect on the adaptability, hence the biology of the chickens. A total of 21 variables available at a 1km by 1km resolution were collected from WorldClim (Hijmans et al., 2005: http://www.worldclim.org/), and the Harmonized World Soil Database v 1.2 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012: http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/). The environmental data included 19 bioclimatic variables and an elevation layer representing current climatic conditions. These 20 layers are commonly used as indicators of annual trends in seasonality, temperature and precipitation. In addition, a land cover layer, total cultivated land, was included as a proxy to anthropogenic intervention and agricultural systems, as smallholders occurrence and poultry density are closely linked in Ethiopia (Dessie, 2003; Mwacharo et al., 2013).

These environmental variables can be correlated. However, to avoid overfitting we used Maxent, which uses a regularization parameter to smoothen the model. It will reduce the importance of variables in the model when they are either of low predictive value or highly correlated to other variables (Phillips et al., 2006). Using this regularization parameter has been shown to perform better than other procedures that use other modelling methods to pre-select variables (Elith and Leathwick, 2009; Elith et al., 2011).

3.2.3 Poultry production system in Ethiopia

Poultry production in Ethiopia is dominated by smallholder producers where nearly all rural and peri-urban families keep a flock of free range chickens in a scavenging system (Moges et al., 2010; Pym, 2010). Village production systems (also denoted traditional or backyard) account for 97% of the poultry production in Ethiopia, making the productivity highly dependent on the environment. In these systems chickens rely almost entirely on scavenging for feed. The amount of nutrients available depends on the region and season of the year. Between rainy seasons feed is limited because the land where chickens usually scavenge is used to grow crops. Attempts to improve the poultry sector in the country have been done through the introduction of exotic chicken breeds, but with no emphasis on changing the husbandry practices. Therefore, exotic breeds are kept under the same backyard conditions as the indigenous chickens (Habte et al., 2017).

3.2.4 Breeds and occurrence

Two exotic breeds were selected for this study based on prior knowledge about their presence in smallholder farms in Ethiopia. The Fayoumi breed originates in Egypt and is said to be adapted to hot and very dry areas in tropical and sub-tropical conditions (Geleta et al., 2013). The Koekoek breed, developed in South Africa, is popular among South African farmers, and said to be adapted to the local conditions in South Africa (Grobbelaar et al., 2010). For Ethiopia, a total of 161 breed locations were used, 62 for the Fayoumi breed, and 99 for the Koekoek breed (Figure 3.1b). These locations were obtained from the national research institute that handles the poultry database in the country; the Ethiopian Institute of Agricultural Research (EIAR).

3.2.5 Predicting breed occurrence

Using the environmental variables selected previously, for each breed independently, we generated a map of the potential distribution given the current climatic and land cover conditions. The range of the potential distributions of both breeds was visualized and assessed in a heat map of the country. To distinguish climatically suitable from unsuitable areas, we applied the “minimum training presence” threshold rule which uses the least suitable training occurrence record as the threshold (Pearson et al., 2007; Norris, 2014). Following the map generation, we validated the model using the receiver operating characteristic (ROC) curve and a binomial test of omission (known areas of presence predicted absent, Phillips et al., 2006). The ROC analysis is a standard approach to test model performance by evaluating the sensitivity (absence of omission error) and 1-specificity (commission error). For each breed the environmental variables that had the highest predictive contribution while building the model were identified.

3.2.6 Cross-validation

To determine if the model predictions could predict breed suitability, we first divided the country in 1 x 1 decimal degree grids, which gave us a total of 110 cells. The grid was applied to limit the effect of spatial clustering on the cross-validation. For each breed independently, instead of removing points one by one, all the localities within each cell where the breed was present were removed from the training data set. This was done cell by cell for all of the cells that included the

occurrence data. Once the occurrence points were removed from the cell, the model was fitted to predict a probability of occurrence for that same cell.
For the cells where the breed was not present, a set of random locations were defined as absent. This set of absent locations was created using ArcGIS v10.3.1 (ESRI, 2015). For each of these cells, the set of random localities were removed from the training data set, and then the model was fitted to estimate the mean predicted probability for each of the cells where the localities were removed. This was done cell by cell for all of the cells with the absence localities. Similar validation designs (variations on the k-fold cross-validation) are used for other approaches in wildlife species to develop more rigorous habitat distribution models (Muscarella et al., 2014).

To base our results on suitable environmental conditions where poultry exists in Ethiopia, we first selected the cells with reported poultry density greater than 10 individuals per square kilometer (Robinson et al., 2014). Then from those cells we selected the ones where each breed was present, hence 78 of the 110 total cells for Fayoumi, and 76 cells of the total 110 cells for Koekoek were included, and the predicted probabilities for the cells with occurrences and the cells with absences were compared. The difference in probabilities between cells with occurrence and with absence was visualized with a density plot; a t-test was applied to test whether both groups were significantly different. An R-script implementing the cross-validation is included.

3.2.7 Breed ranking

For each of the nine regional states in Ethiopia, we established the percentage of area predicted as suitable. Using ArcGIS v10.3.1 (ESRI, 2015) we calculated the total area per region, and the total area predicted as suitable for each of the breeds in each region. Finally for each breed we ranked the five regions that had the highest percentage of potential suitable area.

Data handling and Maxent algorithm (Phillips et al., 2006) which was implemented using the dismo package (Hijmans et al., 2017) , were conducted with R version 3.2.2 (R Development Core Team, 2016) running on RStudio version 0.99.902 (RStudio Team, 2015).

3.3 Results

3.3.1 Prediction and ranking of suitability for breeds

For both breeds, the model predicted that suitable environmental conditions exist beyond the current distribution of the breed (Figure 3.2a, Figure 3.2b). The area under the ROC curve for the model predicting the potential distributions of the Fayoumi and Koekoek breeds was close to one (0.981 and 0.975 respectively), indicating that the model performed well.

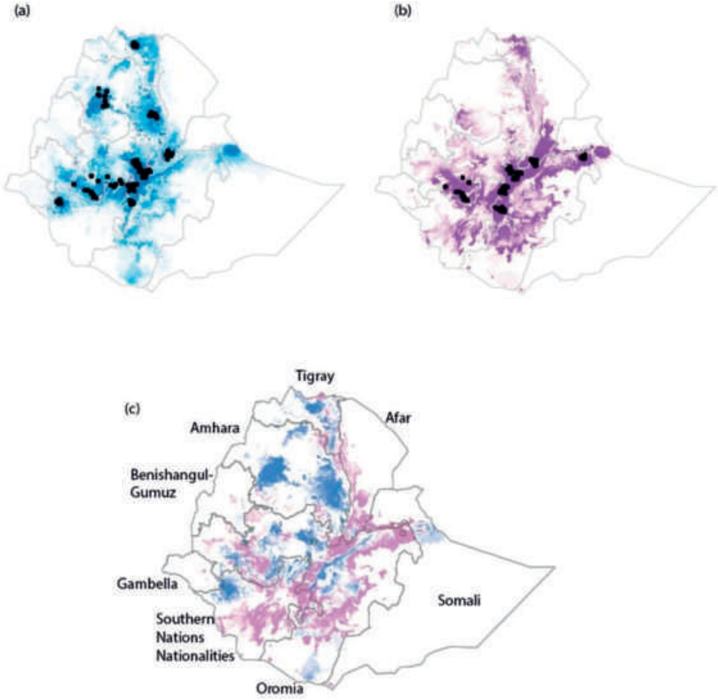

Figure 3.2 Suitability predictions for (a) Koekoek, and (b) Fayoumi breeds in Ethiopia. Predicted areas are shaded; darker colors denote areas of higher climatic suitability. Observed localities used to build the model are shown in black dots. Ratio of suitability between breeds (c). Purple color indicate higher predicted suitability for Fayoumi than for Koekoek. Blue color indicate higher predicted suitability for Koekoek than for Fayoumi.

The percentage of area predicted as suitable for each of the nine regional states in Ethiopia differed between breeds (Figure 3.2c). For the Fayoumi breed, the four regional states with highest percentage of area predicted as suitable were Oromia, Southern Nations Nationalities and People's Region (SNNPR), Amhara and Tigray

(10.9%, 9.13%, 1.29% and 0.57% respectively; Table 3.2). For the Koekoek breed the four regional states with highest percentage of area predicted as suitbable were Amhara, Oromia, Southern Nations Nationalities and People's Region (SNNPR), and Tigray (12.93%, 10.41%, 9.45% and 0.74% respectively; Table 3.2).

Table 3.2 Percentage of area predicted as suitable in the top four regions for each breed.

Chicken breed	Region	Percentage of area predicted as suitable
Koekoek	Amhara	12.93
	Oromia	10.41
	SNNPR	9.45
	Tigray	0.74
Fayoumi	Oromia	10.9
	SNNPR	9.13
	Amhara	1.29
	Tigray	0.57

3.3.2 Most important climatic conditions

Differences in habitat suitability were supported by differences in environmental conditions. The variable explaining most of the variation in suitability for the Fayoumi breed (43.7%; Table 3.2) was associated to total cultivated land. The next two axes (jointly accounting for 26.3% of the environmental variation; Table 3.3) were associated to precipitation. For the Koekoek breed, the variable explaining most of the variation (PC1; 18.1%; Table 3.3) was the minimum temperature of the coldest month, and the next two axes (jointly accounting for 21.6% of the variation. Table 3.3) were associated to mean temperature of the warmest quarter, and the range of mean monthly temperature.

Table 3.3 Selected environmental variables with their percent contributions to the prediction for each breeds' model using Maxent.

Chicken breed	Environmental variable	Percentage contribution
Koekoek	Minimum temperature of coldest month	18.1
	Mean temperature of warmest quarter	11.9
	Mean diurnal range	9.7
Fayoumi	Total cultivated land	43.7
	Precipitation of driest quarter	16.9
	Precipitation of coldest quarter	9.4

3.3.3 Cross-validation

For each of the breeds the mean predicted suitability for the occurrence cells was greater than the mean predicted suitability for the absence cells ($P < 0.05$). For the Koekoek breed, the mean predicted suitability for the cells with absences was 0.047, and for the cells with occurrences was 0.167. For the Fayoumi breed, the mean predicted suitability for the absences was 0.036 and 0.249 for the occurrences (Figure 3.3).

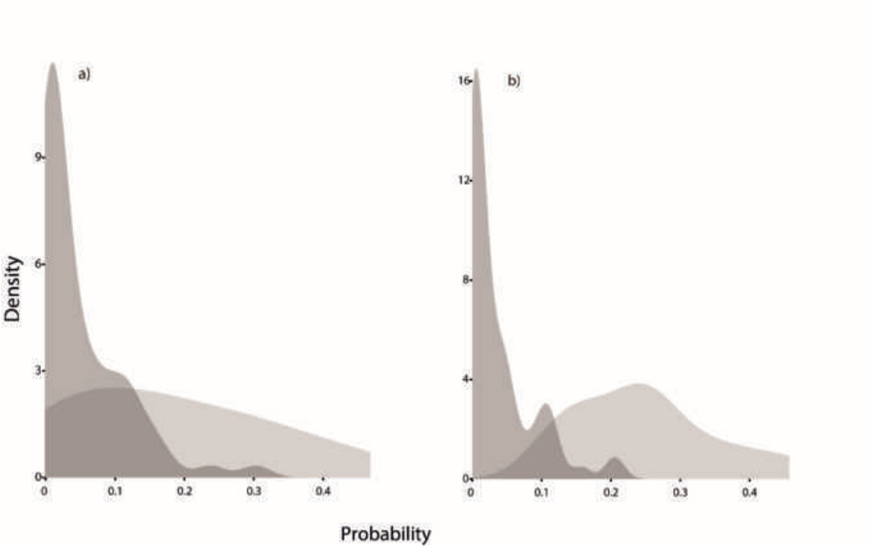

Figure 3.3 Density plots showing the probability predicted as suitable for the cells where the (a) Koekoek and (b) Fayoumi breeds occur (in light grey) and where they are absent (dark grey).

3.4 Discussion

A variety of GIS-based tools have been applied in agriculture. In goats and sheep they have been used to characterize their production system (Malafant, 1998), to propose pasture areas in regions where land has been fragmented (Kalivas and Apostolopoulos, 2005), and to analyze the spatial link between indigenous breeds and areas of livestock usage (Bertaglia et al., 2007). In cattle, buffaloes and sheep, GIS has been applied to see the spatial structure of animal populations, and to evaluate the characteristics of disease transmission between farms (Cringoli et al., 2007). In domestic fowl species, GIS was used to examine the extent of the ecological tolerance of an ancestor bird species to evaluate the success of domestication (Pitt et al., 2016). However, the use of habitat prediction models based on climate and land cover, have not been applied to an animal-breeding context. Here we show how these tools that are widely applied in wild species to cover diverse topics in biogeography, conservation and climate change, can be applied to in livestock to predict breed specific environmental suitability.

Our results suggest that the two breeds that were tested occupy different climatic environments; the Fayoumi breed is suitable for areas where there is a higher percentage of land used in agriculture, and where there is higher precipitation, whereas the Koekoek breed is suitable in colder environments with larger temperature fluctuation. Even though in our dataset the breeds were kept in overlapping areas of the country, they do not always occur together. The Koekoek breed is kept in some localities with tepid to cool moist and sub-moist mid-highlands. The Fayoumi breed is kept in tepid to cool humid mid-highlands, and hot to warm moist lowlands. Temperature and rainfall were found to be the main drivers of the differences in the potential distribution. These climatic parameters are likely to affect livestock production and are highly distinctive between the agro ecologies within our dataset.

The distribution models indicated that the suitable areas for both of the breeds extend beyond their current boundaries, which suggests that there are more areas of the country where the breeds could be suitable for poultry production. The model was sensitive enough to distinguish between breeds. Areas that were predicted as highly suitable differed between the breeds, were found to have significant climatic differences. For the Koekoek breed, the model predicted higher suitable cooler areas in the northern and southern parts of Ethiopia, while for the Fayoumi breed, humid areas towards the center of the country were predicted as highly suitable. Knowledge on the environmental conditions that can have an effect on the breeds' performance is of crucial importance when deciding where to introduce them and where to maintain them.

Adaptability to different environments can be explained by looking at the breeds' origin, where environmental and anthropogenic selective pressures have shaped their adaptation to specific environments. The Fayoumi is a breed of Egyptian origin (Hossaryl and Galal, 1994), while the Koekoek originated in South Africa (Grobbelaar et al., 2010). A study that assessed the genetic diversity of chicken populations in Africa, Asia and Europe revealed that the Fayoumi breed was grouped with chickens from the Mediterranean, while the Koekoek shared a cluster with eastern European breeds and broiler chickens (Lyimo et al., 2014). This genetic origin suggests that breeds might respond distinctively in different agro-ecologies.

Even though the origin of the breeds was not in Ethiopia, we interpreted its current occupation area as a success in productivity and as evidence for suitability in the current range. Therefore, the current area of occupation could be used to predict suitability for other regions in the country where the breeds are not present.

This novel approach can find practical use in breeding programs, as it can be applied at different scales for different livestock breeds. For region specific breeds, such as the indigenous Horro chickens (Wondmeneh, 2015), or cosmopolitan breeds, such as the Holstein Friesian cattle, these tools can be useful to predict suitability to a given region, given the climatic variables. The approach can be used when the interest is in designing a breeding plan, introducing a breed to a new area, or when trying to understand differences in performance within the same breed in different areas, or between breeds in the same area. To extend the use of prediction models, further analysis can be explored by taking productivity data into account. However, productivity data is difficult to obtain from smallholder farms.

Understanding the environmental requirements of different breeds is an important tool to support higher productivity in particular regions (Arthur and Albers, 2003). As regions can have different environmental conditions, it is imperative to understand how livestock adapt to their environment, and which variables are shaping the differences in performance between breeds.

Breeding programs in developing countries are often ineffective as a consequence of the non-adaptability of the introduced breeds to the challenging environments (Montaldo, 2001; Ojango and Pollott, 2002). More recently Ferreira et al. (2017) and Rosé et al. (2017) showed that differences between temperate and tropical climates can cause significant G x E which affects productivity. This breed-by-environmental mismatch is usually estimated as genotype by environment interaction (G x E), the genetic correlation for a given set of traits estimated in two environments. Given the genetic correlation, our methodology can be used to analyze these two environments and predict in which regions a breed will most likely exhibit an environmental mismatch. By analogy it can also reveal potential areas of successful introduction, contributing to a successful breeding program.

In conclusion, this work showed the utility of habitat distribution models applied to a livestock research. This allows making predictions of breed specific suitability taking into account environmental information. Being able to explain the role of genotype by environment interaction can be a useful application of the methodology developed here, that will further help in providing support when designing breeding programs, or introduction programs for local animal production, by understanding the environmental variables that can have an impact on breed productivity between environments.

3.5 Acknowledgements

We sincerely thank the Koepon Foundation for providing a scholarship to the first author. We would also like to thank E. Wondmeneh and staff at the EIAR for providing data.

References

Arthur, J. A., and G. A. Albers. 2003. Industrial perspective on problems and issues associated with poultry breeding. Poultry genetics, breeding and biotechnology: 1-12.

Bertaglia, M., S. Joost, and J. Roosen. 2007. Identifying European marginal areas in the context of local sheep and goat breeds conservation: A geographic information system approach. Agricultural Systems 94: 657-670.

Cringoli, G. et al. 2007. Geo-referencing livestock farms as tool for studying cystic echinococcosis epidemiology in cattle and water buffaloes from southern Italy. Geospatial health 2: 105-111.

Deressa, T. T., C. Ringler, and R. M. Hassan. 2010. Factors affecting the choices of coping strategies for climate extremes. The case of farmers in the Nile Basin of Ethiopia IFPRI Discussion Paper 1032.

Dessie, T. 2003. Phenotypic and genetic characterization of local chicken ecotypes in Ethiopia. Dessertation for Award of PhD degree at Humboldt-University, Berlin, Germany.

Elith, J. et al. 2006. Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. Ecography 29: 129-151.

Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics 40: 677-697.

Elith, J. et al. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43-57.

ESRI. 2015. ArcGIS. Version 10.3.1., Redlands, CA: Environmental Systems Research Institute.

FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonized World Soil Database (version 1.2), FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Ferreira, J. L. et al. 2017. Climate spatialization and genotype-environment interaction effects on weaning weights of Nellore Cattle in extensive systems in tropical regions of Brazil. Ciência Animal Brasileira 18.

Geleta, T., S. Leta, and E. Bekana. 2013. Production performance of Fayoumi chickens under intensive management condition of Adami Tulu research center. International Journal of Livestock production 4: 172-176.

Grobbelaar, J., B. Sutherland, and N. Molalakgotla. 2010. Egg production potentials of certain indigenous chicken breeds from South Africa. Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales 46: 25-32.

Habte, T. et al. 2017. Guide to chicken health and management in Ethiopia: For farmers and development agents. ILRI.

Haftu Kebede Sebho. 2016. Exotic Chicken Status, Production Performance and Constraints in Ethiopia: A Review. Asian Journal of Poultry Science, 10: 30-39.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hijmans, R. J., and C. H. Graham. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global change biology 12: 2272-2281.

Hijmans, R. J., S. J. Phillips, J. R. Leathwick, and J. Elith. 2017. Species distribution modeling. R package version 1.1-4.

Hossaryl, M., and E. Galal. 1994. Improvement and adaptation of the Fayoumi chicken. Animal Genetic Resources Information 14: 33-39.

Kalivas, D., and C. Apostolopoulos. 2005. The use of GIS to improve the resources utilisation of mountain areas: the case of sheep and goat breeding in the Greek regions of Thessaly and Epirus. The European Association for Animal Production 115: 466.

Köhler-Rollefson, I., H. S. Rathore, and E. Mathias. 2009. Local breeds, livelihoods and livestock keepers' rights in South Asia. Tropical animal health and production 41: 1061.

Kosgey, I., R. Baker, H. Udo, and J. Van Arendonk. 2006. Successes and failures of small ruminant breeding programmes in the tropics: a review. Small Ruminant Research 61: 13-28.

Lyimo, C. et al. 2014. Global diversity and genetic contributions of chicken populations from African, Asian and European regions. Animal genetics 45: 836-848.

Malafant, K. 1998. Mapping Livestock Populations, http://www.complexia.com.au/Documents/Density_maps/stock.html.

Mengistu, A. 2003. Country Pasture/Forage Resource Profiles, Ethiopia. Available at: www.fao.org/AG/agp/agpc/doc/Counprof/Ethiopia/Ethiopia.htm#7.%20RESEARCH%20AND%20DEVELOPMENT%20ORGANIzATIONS%20AND (accessed February 2017).

Mirkena, T. et al. 2010. Genetics of adaptation in domestic farm animals: A review. Livestock Science 132: 1-12.

Moges, F., A. Mellesse, and T. Dessie. 2010. Assessment of village chicken production system and evaluation of the productive and reproductive performance of local chicken ecotype in Bure district, North West Ethiopia.

Montaldo, H. H. 2001. Genotype by environment interactions in livestock breeding programs: a review. Interciencia 26: 229-235.

Muscarella, R. et al. 2014. ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecology and Evolution 5: 1198-1205.

Mwacharo, J. M., G. Bjørnstad, J. L. Han, and O. Hanotte. 2013. The History of African Village Chickens: an Archaeological and Molecular Perspective. African Archaeological Review 30: 97-114.

Norris, D. 2014. Model thresholds are more important than presence location type: Understanding the distribution of lowland tapir (*Tapirus terrestris*) in a continuous Atlantic forest of southeast Brazil. Tropical Conservation Science 7: 529-547.

Ojango, J. M. K., and G. E. Pollott. 2002. The relationship between Holstein bull breeding values for milk yield derived in both the UK and Kenya. Livestock Production Science 74: 1-12.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global ecology and biogeography 12: 361-371.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. Townsend Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of biogeography 34: 102-117.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological modelling 190: 231-259.

Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161-175.

Pitt, J., P. K. Gillingham, M. Maltby, and J. R. Stewart. 2016. New perspectives on the ecology of early domestic fowl: An interdisciplinary approach. Journal of Archaeological Science 74: 1-10.

Pym, R. 2010. Poultry feed availability and nutrition in developing countries Food and Agriculture Organization of the United Nations, Poultry Development Review.

R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved on 8 December 2016 from http://www.R-project.org.

Robinson, T. P. et al. 2014. Mapping the Global Distribution of Livestock. PLoS ONE 9: e96084.

Rosé, R. et al. 2017. Interactions between sire family and production environment (temperate vs. tropical) on performance and thermoregulation responses in growing pigs. Journal of Animal Science 95: 4738-4751.

RStudio Team. 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, USA. Retrieved on 30 June 2015 from http://www.rstudio.com/.

Soberón, J., and M. Nakamura. 2009. Niches and distributional areas: concepts, methods, and assumptions. Proceedings of the National Academy of Sciences 106: 19644-19650.

Solti, L., E. Crichton, N. Loskutoff, and S. Cseh. 2000. Economical and ecological importance of indigenous livestock and the application of assisted reroduction to their preservation. Theriogenology 53: 149-162.

Wondmeneh, E. 2015. Genetic Improvement in Indigenous Chicken of Ethiopia (PhD Thesis) Wageningen Univeristy, Wageningen, The Netherlands.

Wurzinger, M., J. Sölkner, and L. Iñiguez. 2011. Important aspects and limitations in considering community-based breeding programs for low-input smallholder livestock systems. Small Ruminant Research 98: 170-175.

Chapter 4

Using phenotypic distribution models to predict livestock performance

Maria Lozano-Jaramillo[1], Setegn W Alemu[2], Tadelle Dessie[2], Hans Komen[1], John WM Bastiaansen[1]

[1]Wageningen University & Research Animal Breeding and Genomics, P. O. Box 338, 6700 AH Wageningen, The Netherlands; [2]International Livestock Research Institute, P. O. Box 5689 Addis Ababa, Ethiopia

Submitted

Abstract

Livestock production systems of the developing world use indigenous breeds that locally adapted to specific agro-ecologies. Introducing commercial breeds usually results in lower productivity than expected, as a result of unfavourable genotype by environment interaction. It is difficult to predict of how these commercial breeds will perform in different conditions encountered in e.g. sub-Saharan Africa. Here, we present a novel methodology to model performance, by using growth data from different chicken breeds that were tested in Ethiopia. The suitability of these commercial breeds was tested by predicting the response of body weight as a function of the environment across Ethiopia. Phenotype distribution models were built using machine learning algorithms, to make predictions of weight in the local environmental conditions based on the productivity for the breed. Based on the predicted body weight, breeds were assigned as being most suitable in a given agro-ecology or region. We identified the most important environmental variables that explained the variation in body weight across agro-ecologies for each of the breeds. Our results highlight the importance of acknowledging the role of environment in predicting productivity in scavenging chicken production systems. The use of phenotype distribution models in livestock breeding is recommended to develop breeds that will better fit in their intended production environment.

4.1 Introduction

In livestock production systems of the developing world, native breeds are known to be locally adapted to specific environmental conditions (Dessie, 2003; Pym, 2010a). This adaptability allows these breeds to fulfil a key role in providing nutrition and income to the rural and peri-urban households, particularly under low input management systems. However, these local breeds exhibit low productivity when compared to exotic breeds (Bekele et al., 2009). As a means to improve local meat and egg production, the African continent has witnessed the introduction of exotic chicken breeds (Mwacharo et al., 2013). Nevertheless, after more than 40 years of exotic breed introduction most attempts have been unsuccessful mainly because of the non-adaptability of these breeds to the local environmental conditions, and the prevailing low input production systems (Dessie et al., 2000; Bekele et al., 2010; Dana et al., 2010b).

Nowadays, the poultry industry in most developing countries is divided in three systems, village, semi-intensive and intensive. Distinctions between these production systems are based on the breeds (indigenous or introduced), but also on the management conditions (Habte et al., 2017). Generally, introduced commercial breeds are kept under better housing and feeding conditions in the semi-intensive and intensive systems, while smallholders keep their indigenous breeds under semi-free range, scavenging feeding conditions in the village systems (Pym, 2010a, 2010b; Habte et al., 2017). Birds in the village systems rely on foraging for water, house wastes, insects, snails, earthworms, and other resources where the availability is highly dependent on the local environment and season (Dessie, 2003; Habte et al., 2017).

There are a few commercial breeds that have been specifically developed for production in (semi-) tropical conditions. Examples include the Kuroiler, a hybrid chicken from India (Sharma et al., 2015), Sasso, a breed originating from France (SASSO Breeding Company, 2018), and the Koekoek, a breed developed in South Africa (Grobbelaar et al., 2010). These breeds are developed to be successful and more productive in low maintenance, tropical systems. However, there is a lack of understanding of how these tropically adapted commercial breeds will perform across the wide range of environmental conditions encountered in sub-Saharan Africa and under diverse smallholder systems.

In a recent study, we used predictive habitat distribution models on the Koekoek breed, to predict areas of potential suitability for this breed in Ethiopia. We also identified which environmental conditions explain the breeds' distribution in the country (Lozano-Jaramillo et al., 2018). For that study no information on the productivity of the breed was taken into account, as only information about the location of the breed was available.

An approach to take the productivity into account, is the use of phenotype distribution models. These models were recently introduced in ecology studies to capture the response of phenotypic traits as a function of environmental conditions (Smith et al., 2017). As the environment plays an important role in scavenging systems, this model can also have value to predict the productivity of a breed in a specific region. The application of phenotypic distribution models to understand the role of the environment in livestock productivity is novel and has not yet been implemented.

Due to the influence of environmental factors on animal performance, research on the mechanisms of livestock adaptation to challenging environments is urgently needed (Van Marle-Köster and Casey, 2001). The African Chicken Genetic Gains project (ACGG; https://africacgg.net/) is testing the performance of different chicken breeds in smallholders' households. ACGG is an Africa-wide collaboration research program led by the International Livestock Research Institute (ILRI) and Bill and Melinda Gates Foundation funded, that aims to improve chicken productivity to benefit rural smallholder households. Knowledge on how the productivity of these introduced chicken breeds is affected by the environment may allow predictions on how they would respond to specific agro-ecologies, and to tailor breeding programs to collect performance data in the most informative environments.

Our main objective was to predict the response of productivity traits of different introduced breeds, as a function of the environment where they were introduced. Using performance data on five chicken breeds from the ACGG project in Ethiopia, the association with environmental parameters was established using phenotypic distribution models. Predictions for productivity traits were made to establish suitable areas for breed distribution. Such predictions could provide insightful information for breeding companies or farmers to decide which areas can be suitable to introduce a breed, or which breed will be better suited for a specific environment.

4.2 Materials and methods

Live body weight data from the five breeds tested in Ethiopia was obtained from the ILRI datasets portal. For details on ACGG see https://africacgg.net/. Environmental variables that characterize the different agro-ecological zones in Ethiopia were obtained from WorldClim (Hijmans et al., 2005) and the Harmonized World Soil Database v 1.2 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). All data cleaning and analyses were undertaken in R version 3.5.1 (R Development Core Team, 2016) running on RStudio version 1.1.383 (RStudio Team, 2015). The experimental protocols and ethical guidelines were approved by the ILRI Institutional Research Ethics Committee (ILRI IREC) under reference number ILRI-IREC2015-08. ILRI IREC is accredited by the National Commission for Science, Technology and Innovation (NACOSTI) in Kenya.

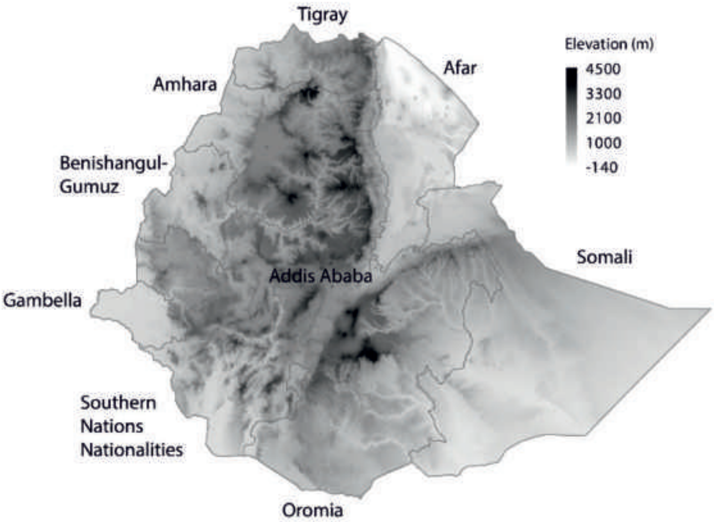

Figure 4.1 Map of Ethiopia showing its regional states.

4.2.1 Chicken breeds

The ACGG program identified four exotic chicken breeds presumably tolerant to sub-Saharan African environmental conditions (Koekoek, Kuroiler, Sasso, and Sasso-RIR), and in 2016 fertile eggs were imported into Ethiopia. A locally improved breed named Horro was also used in parallel to the exotic breeds. The latter one assumed to be locally adapted, and genetically improved for growth and egg production in the environmental conditions of the low input systems characterizing the smallholders farms in Ethiopia (Wondmeneh, 2015).

The five breeds were tested in 63 villages, divided over five regions across Ethiopia (Addis Ababa, Amhara, Oromia, Southern Nations Nationalities, and Tigray; see Figure 1). A total of 1393 households received approximately 25 six-week old chicks per household. All breeds were tested in all the villages, but only one breed was tested in each household. The introduced chickens were exposed to the same environmental and management conditions as the local breeds already kept by the farmers. The breed distribution started in August 2016, and data collection ended in January 2018.

4.2.2 Breed and phenotypic data

Households were georeferenced, and each breed was randomly assigned to a household. Chickens were delivered at 6 weeks of age. Traditionally in Ethiopia, farmers sell the male chickens when they reach 2 kg, which is around 20 weeks. For male chickens, weight data was collected until 20 weeks of age. Females are kept longer for egg production, therefore, weight data for females was collected until 72 weeks of age.

4.2.3 Data cleaning

In each household, body weight data was collected every two weeks as a group measure. Taking into account the number of birds weighted, we converted the group weight to average individual body weight. After calculating the average individual weight, we first eliminated all observations with weight less than 50 grams. We deleted the households with more than 30 birds reported per household. We also deleted those observations where age was less than 6 weeks. For males we used male body weight during the growing phase (weeks 14-19) and for females we divided the data into growing phase (weeks 14-19), and adult phase (weeks 20-72). Not all the weeks had data for all of the households. Therefore, we standardized the body weight in each household to the average value of week in each age phase for all breeds for males and females independently. The average week value for the growing phase for males and females was 16.54 and 16.52 respectively. For females during the adult phase this was 41.81. Then, the following linear model was fitted:

$$\boldsymbol{y} = \mathbf{H} + \mathbf{W} + \boldsymbol{e} \tag{4.1}$$

where $\boldsymbol{y}$ is the vector with the average bird weights, **H** is the household at the week of measurement, **W** is the week of measurement, and e is the vector containing the random residual error. We assumed that within breeds all chickens had the same body weight at the time of hatching and fitted a model with common intercept and different slope per household. For each household for each age phase independently, we calculated the least squares means (LSmeans), using the emmeans package in R (Lenth, 2019) .

Table 4.1. Environmental variables used to build the models.

Variables	**Source**
Annual Mean Temperature	WorldClim
Mean Diurnal Range (Mean of monthly (max temp - min temp))	WorldClim
Isothermality (BIO2/BIO7) (* 100)	WorldClim
Temperature Seasonality (standard deviation *100)	WorldClim
Max Temperature of Warmest Month	WorldClim
Min Temperature of Coldest Month	WorldClim
Temperature Annual Range (BIO5-BIO6)	WorldClim
Mean Temperature of Wettest Quarter	WorldClim
Mean Temperature of Driest Quarter	WorldClim
Mean Temperature of Warmest Quarter	WorldClim
Mean Temperature of Coldest Quarter	WorldClim
Annual Precipitation	WorldClim
Precipitation of Wettest Month	WorldClim
Precipitation of Driest Month	WorldClim
Precipitation Seasonality (Coefficient of Variation)	WorldClim
Precipitation of Wettest Quarter	WorldClim
Precipitation of Driest Quarter	WorldClim
Precipitation of Warmest Quarter	WorldClim
Precipitation of Coldest Quarter	WorldClim
Elevation	WorldClim
Total cultivated land	Harmonized World Soil

4.2.4 Environmental data

The introduced chickens were kept in low input systems along with indigenous chicken. Some feed supplementation was given by farmers at their discretion. The supplemented feed was either grown by the farmers or bought externally, the rest of the feed intake was from scavenging. The introduced chickens had to cope with the same environmental conditions as the indigenous chickens, where food availability depends on seasonality. Therefore, we used a total of 21 variables at a 1km by 1km resolution (Table 4.1). The environmental data, representing current conditions included 19 bioclimatic variables and an elevation variable. These 20 variables represent trends in temperature, seasonality and precipitation, variables that are expected to have an influence on the chicken productivity and food availability. Smallholders in Ethiopia tend to have crops in their households, which has an influence on what is fed to their livestock. To represent this link between human intervention and agriculture (Dessie, 2003; Mwacharo et al., 2013) we used an additional layer, the total cultivated land. This variable represents the, that the percentage of area that is used in agriculture, therefore having an impact on the food availability.

4.2.5 Phenotypic variation models

In ecology, regression analyses are often used to explain the relationship between ecological data and to build prediction models. Ecological data can sometimes be complex showing non-linear relationships and/or spatial and temporal correlation. To address these problems, more flexible, but still complex models are often used, such as generalized additive models (GAMs). GAMs can identify nonlinear approaches, and still generate easy to interpret relationships. However, including too many covariates will lead to overfitting, which decreases the prediction accuracy. Therefore, several machine learning algorithms, such as boosting, have been used to increase the prediction accuracy of standard regression models (Maloney et al., 2012).

To model the relationship between environmental variables and chicken phenotypes, the values of the 21 environmental variables for each of the georeferenced household locations were used as predictor variables. The response variable was the LSmeans for body weight per household for each of the breeds at the average week for each growing phase. To increase prediction accuracy, gradient boosting, a machine learning algorithm, was applied to a GAM. Gradient boosting is an iterative process, where models are built in a step-wise fashion, starting from weak prediction models that fit the data poorly (base-learners), and by learning from the previous step, the fitting is improved in the next iteration (Maloney et al., 2012). All the models were built using all the environmental variables as base-learners. Boosting algorithms use as tuning parameter a stopping iteration, which determines the optimal point of where the algorithm should stop before convergence (e.i. early stopping). This choice of stopping iteration becomes crucial, as it prevents overfitting the data and improves the accuracy. The model for each breed was run independently, and a separate stopping iteration was determined for each model using the method described in Mayr et al. (2012). The predicted body weight was represented in a heatmap that covers all the regions of the country were chickens are known to be kept. Desert areas (Somali and Afar regions (see Figure 1) were excluded from the predictions, as the environmental conditions in these regions lay outside the range of the environmental conditions in which the chickens where tested (Supplementary Tables S4.1 and S4.2).

The variable of highest contribution in building each of the models was obtained. For this, the in-bag risk reductions per boosting step of a fitted model are accumulated individually for each variable (base-learner) contained in the model. This quantifies the individual variable contribution to risk reduction of each variable (base-learner), and can be used to compare the importance of different variable in the model (Hothorn et al., 2018). To quantify the precision of our predictions, we estimated the correlation between the predicted value obtained from the phenotypic distribution models, and the LSmeans per household in each region. The LSmeans are the observed values for each of the households. With the model we predict the performance. If the predictions are precise, then the correlation between the predictions and the LSmeans would be high. Data cleaning, modelling the phenotypic variation and importance variable selection were done using the package mboost in R (Hothorn et al., 2018).

4.3 Results

4.3.1 Estimated body weights

LSmeans for male body weights were estimated for week 16.54 (growing phase; Table 4.2). The heaviest breed was the Sasso with a mean estimated body weight of 1035.7g, and the lightest was the Horro with a mean of 635.3g. LSmeans for female body weights were estimated for week 16.52 (Table 4.3). The heaviest breed was the Koekoek with a mean estimated body weight of 1280.2g, and the lightest was the Horro with a mean estimated weight of 738.6g. LSmeans for female body weights in week 41.81 (adult phase) were highest for the Sasso-RIR with a mean weight of 2863.3g, and the lightest remained to be the Horro with a mean estimated body weight of 2250.8g (Table 4.4).

Table 4.2 Minimum, maximum and mean estimated (LSmeans) male body weight for each of the five breeds across Ethiopia during the growing phase (week 16.54).

Breed	Minimum estimated body weight (g)	Maximum estimated body weight (g)	Mean estimated body weight (g) (S.E)
Horro	403.9	1119.2	635.3 (12.3)
Koekoek	401.9	1675.4	842.5 (11.2)
Kuroiler	436.8	1678.3	919.9 (16.3)
Sasso	504.3	2173.2	1035.7 (21.2)
Sasso-RIR	398.4	1554.8	745.6 (11.6)

4.3.2 Predicted body weights

Predicted body weights varied considerably between breeds across Ethiopia. In general, for the growing phase of males and females, the models predicted for the Sasso breed to have the highest body weight with 2185.2g for males and 3862.8g for females (Figure 4.2; Table 4.2 and 4.5). The Horro breed is predicted to have the lowest body weight with 1327g for males and 2540.3g for females (Figure 4.3; Table 4.5 and 4.6). The female body weight predicted for the adult phase was the highest for the Sasso breed with 4330.8g. The lightest breed predicted was the Horro with 3074.6g (Figure 4.4; Table 4.7).

Table 4.3 Minimum, maximum and mean estimated (LSmeans) female body weight for each of the five breeds across Ethiopia during the growing phase (week 16.52).

Breed	Minimum estimated body weight (g)	Maximum estimated body weight (g)	Mean estimated body weight (g) (S.E)
Horro	515.6	1058.1	738.6 (15.9)
Koekoek	534.4	1658.6	1280.2 (10.5)
Kuroiler	565.5	1227.3	896.3 (17.3)
Sasso	569.9	1415.2	1081.2 (12.7)
Sasso-RIR	512.4	1500.9	1041.1 (11.3)

To assess the precision of our predictions, correlations between the predicted values from the phenotypic variation models and the household LSmeans were estimated for each growing phase and breed within each of the regions. There is considerable variation within region for all breeds and growing phases, both in estimated and predicted weights (Supplementary Figure S4.1-S4.3). The correlation estimates have a median value of 0.61 but vary by breed and region ranging from 0.18 to 0.89. As suggested by Bonett and Wright (2000), to better estimate a correlation, the sample size should be greater than 25. Therefore, the correlations were only estimated for those regions where the total number of households for each breed was greater than 25. For the lowest number of households, N = 28, the S.Es were 0.098 and 0.103. For the highest number of household, N = 155, the S.E was 0.016 (Supplementary Table S4.3; Supplementary Figure S4.1-S4.3).

Table 4.4 Minimum, maximum and mean estimated (LSmeans) female body weight for each of the five breeds across Ethiopia during the adult phase (week 41.81).

Breed	Minimum estimated body weight (g)	Maximum estimated body weight (g)	Mean estimated body weight (g) (S.E)
Horro	1323.1	2910.8	2250.8 (8.8)
Koekoek	1961.4	3149.8	2703.9 (10.7)
Kuroiler	1565.5	3281.3	2796.1 (9.2)
Sasso	1726.0	3307.8	2821.2 (10.1)
Sasso-RIR	1995.4	3399.6	2863.3 (7.6)

4.3.3 Region suitability

Predictions of optimal region per breed, and the most important variable in building the model are given in Table 4.5, 4.6 and 4.7. Results for male and female body weights during the growth phase and adult weights were very similar. Sasso always had the highest predicted body weights in Tigray. Koekoek had the highest weights in Amhara, while Kuroiler had the highest weights in Oromia. For Horro and Sasso-RIR, the pattern was less clear. Horro was predicted to perform best in either Amhara or Tigray for males and females. For Sasso-RIR there was no clear single region where the breed was expected to perform best for all sex and age groups.

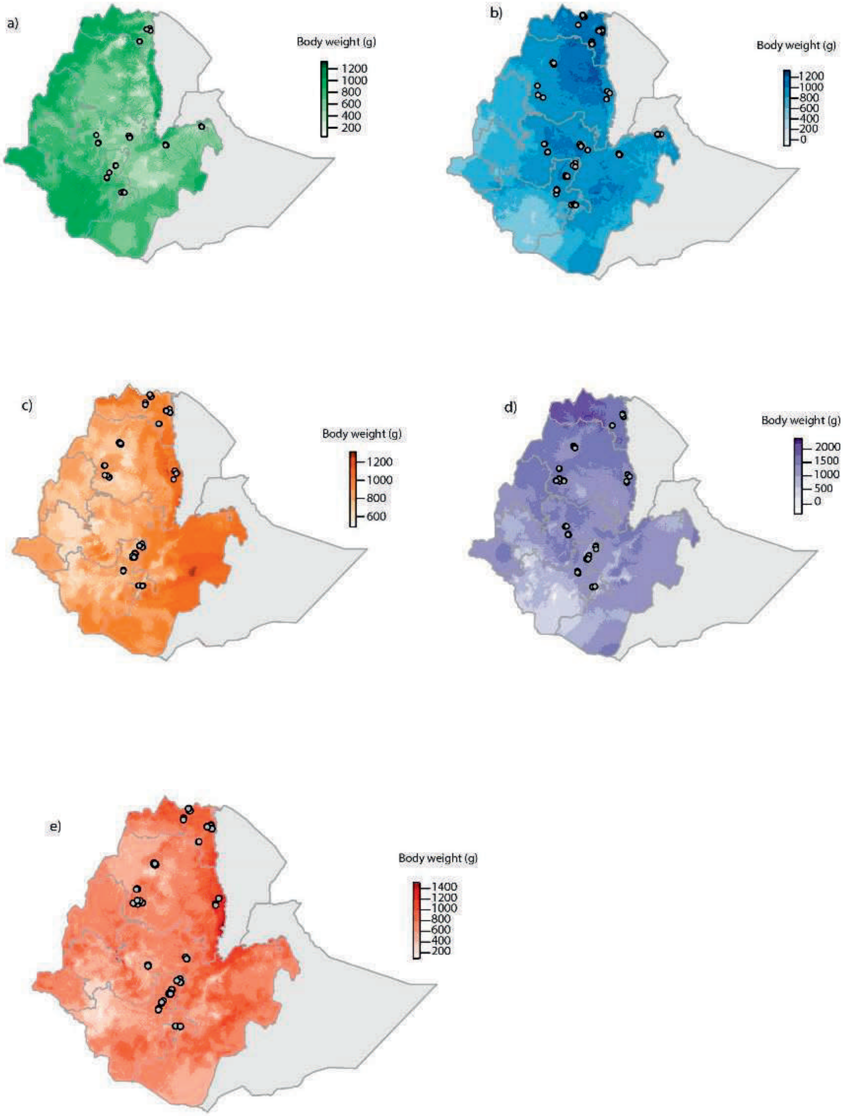

Figure 4.2 Predicted current phenotypic measures for male body weight in grams during the growing period (weeks 14 to 19) for a) Horro, b) Koekoek, c) Kuroiler, d) Sasso, and e) Sasso-RIR. Circles in the maps denote the household locations where the phenotypic data was collected. Color scales in each of the maps reflect the predicted body weight, where darker colors indicate higher live weight predicted.

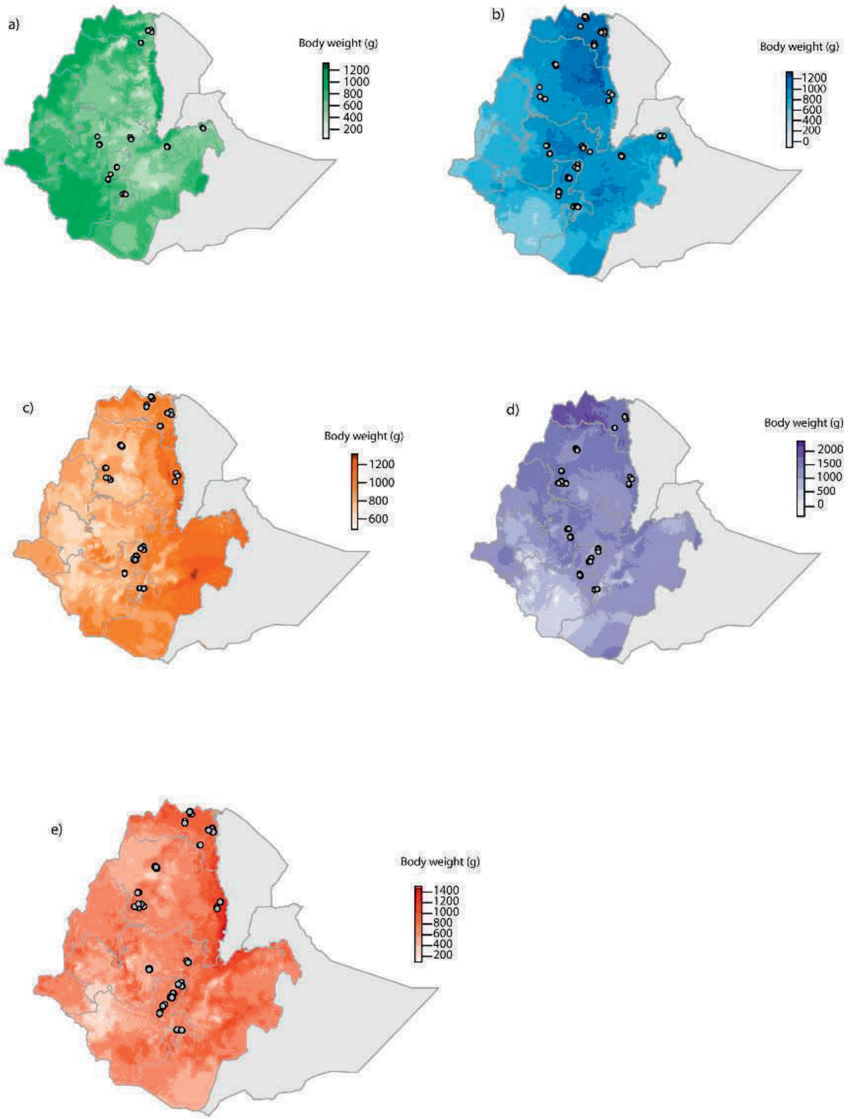

Figure 4.3 Predicted current phenotypic measures for female body weight in grams during the growing period (weeks 14 to 19) for a) Horro, b) Koekoek, c) Kuroiler, d) Sasso, and e) Sasso-RIR. Circles in the maps denote the household locations where the phenotypic data was collected. Color scales in each of the maps reflect the predicted body weight, where darker colors indicate higher live weight predicted.

4.3.4 Variables contributing to the models

The individual contribution from each environmental variable to build the best model were identified. The environmental variables of strongest influence for the predicted male body weight distribution were different for each breed. Patterns of male body weights for Horro, Koekoek, Sasso and Sasso-RIR were best predicted by variables associated with temperature. For Kuroiler, precipitation of the coldest quarter was the variable that had the strongest influence on the distribution (Table 4.5). For female body weights during the growing phase, the patterns of the body weight distribution for Horro and Sasso were again best predicted by temperature related variables. However, for Koekoek, Kuroiler and Sasso-RIR, precipitation variables had more influence on the distribution (Table 4.6). For female body weights during the adult phase, the body weights distribution patterns for Horro, Koekoek and Kuroiler were best predicted by variables associated with precipitation. For Sasso and Sasso-RIR, temperature annual range and elevation respectively had the largest influence on the body weights distribution (Table 4.7).

Table 4.5 Minimum and maximum male body weights predicted for growing phase. The most important environmental variable for building the model and region predicted for higher body weights is shown per breed.

Breed	Minimum predicted weight (g)	Maximum predicted weight (g)	Environmental variable of importance	Region with higher body weights predicted
Horro	216.8	1327	Max. temp. or warmest month	Tigray
Koekoek	161.1	1381.3	Mean diurnal range	Amhara
Kuroiler	656.4	1448.8	Precipitation of coldest quarter	Oromia
Sasso	282	2185.2	Mean diurnal range	Tigray
Sasso-RIR	270.5	1558.6	Isothermality	Amhara

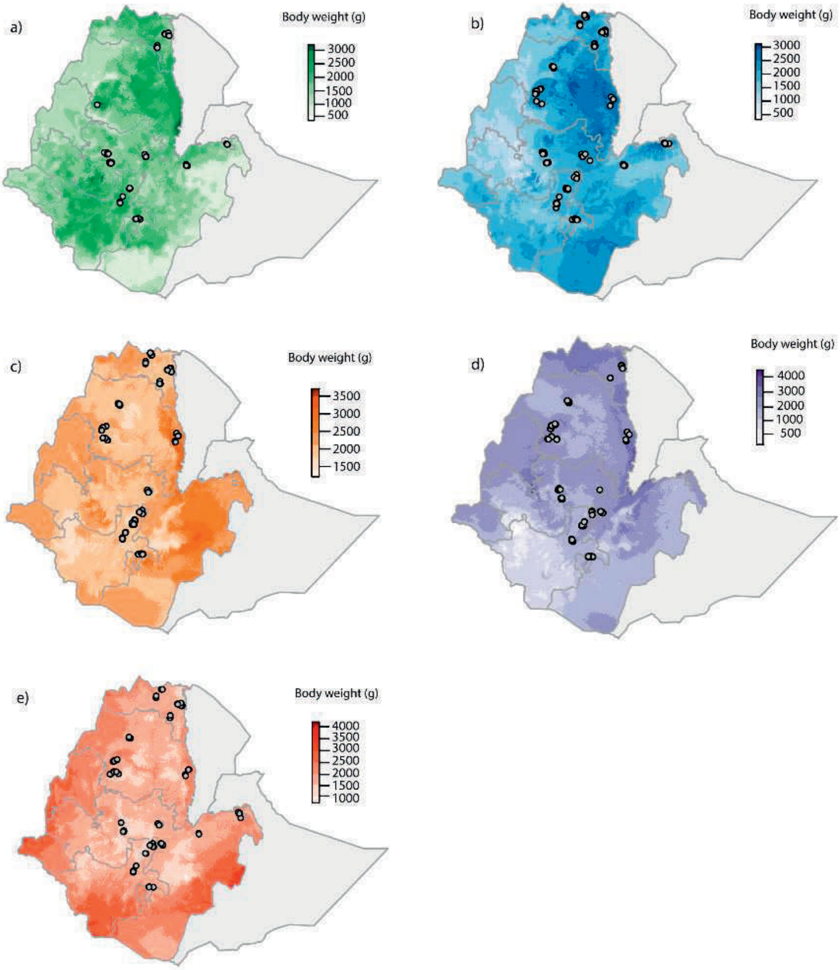

Figure 4.4 Predicted current phenotypic measures for female body weight in grams during the adult period (weeks 20 to 72) for a) Horro, b) Koekoek, c) Kuroiler, d) Sasso, and e) Sasso-RIR. Circles in the maps denote the household locations where the phenotypic data was collected. Color scales in each of the maps reflect the predicted body weight, where darker colors indicate higher live weight predicted.

Table 4.6 Minimum and maximum female body weights predicted for growing phase. The most important environmental variable for building the model and region predicted for higher body weights is shown per breed.

Breed	Minimum predicted weight (g)	Maximum predicted weight (g)	Environmental variable of importance	Region with higher body weights predicted
Horro	292.3	2540.3	Temperature annual range	Tigray
Koekoek	186.1	2635.5	Precipitation of wettest month	Amhara
Kuroiler	1022	3545.1	Precipitation of wettest month	Oromia
Sasso	524.6	3862.8	Annual temperature range	Tigray
Sasso-RIR	341.1	2556.8	Precipitation of driest month	Oromia

Table 4.7 Minimum and maximum female body weights predicted for adult phase. The most important environmental variable for building the model and region predicted for higher body weights is shown per breed.

Breed	Minimum predicted weight (g)	Maximum predicted weight (g)	Environmental variable of importance	Region with higher body weights predicted
Horro	388.1	3074.6	Precipitation of driest month	Amhara
Koekoek	225.1	3344.5	Precipitation of warmest quarter	Amhara
Kuroiler	1107.5	3746.2	Precipitation of warmest quarter	Oromia
Sasso	635.7	4330.8	Temperature annual range	Tigray
Sasso-RIR	730	4154.6	Elevation	Gambella

4.4 Discussion

We used phenotypic distribution models, implementing gradient boosting, a machine learning technique, to increase the predictive ability of GAMs. Machine learning procedures improve the accuracy of the models, by addressing overfitting, and variable selection simultaneouslyBroadly, our results demonstrate the importance of understanding how different breeds respond to environmental conditions, and how different environmental variables can influence their productivity.

In Ethiopia, the Horro breed was characterized and genetically improved for growth traits (Dana et al., 2010b; Wondmeneh, 2015). The breeding program was conducted under an intensive management system, including high quality poultry houses, formulated feed, and vaccination. The ACGG program is the first attempt to test the ability of this breed to adapt to on-farm environment and management conditions (Wondmeneh, 2015). Even though our results show higher body weights than previously reported for the breed (Dana et al., 2010a), compared to the introduced commercial lines they exhibit lower growth performance. The Horro breed originates in the Oromia region, at the Horro district at altitudes ranging between 2580 to 2810 meters above sea level (masl) in cool wet highlands. The average annual temperature is 13.3°C, and the seasons can be divided in three; a main rainy season, a dry season, and a short rainy season. The seasonal variety in the Horro district can explain why the Horro breed seems to perform better in low humidity areas with low levels of precipitation. A recent study showed that during wet periods, the Horro chickens are prone to die due to disease, and survive better in dry seasons (Bettridge et al., 2018).

The Sasso breed originates from the south of France in warm, and dry areas. In accordance to what we find here, we see that the Sassos' performance across ages is linked to temperature. Different studies have shown, in line with our findings, that the Sasso breed was more productive than the indigenous breeds in Ethiopia (Aman et al., 2017). We identified that temperature-associated variables have an influence on the Sasso predicted body weight distribution during all the age phases evaluated. The Sasso-RIR tested in Ethiopia was obtained through a private poultry farm, were crosses were made to generate the breed for the ACGG program (Aman et al., 2017). Therefore, no other studies have been reported using the Sasso-RIR cross to evaluate performance under scavenging conditions. For Sasso-RIR, different environmental variables are responsible for shaping the predicted body weight distribution depending on the age phase, suggesting that the breeds response to the agro-ecology depends on age. In comparison, the pure breed Sasso, was predicted to be heavier than the crossbred Sasso-RIR for the male and female growing periods. For the adult females, the prediction was similar for both breeds. It should be noted that the body weight patterns and the variables of importance in building the models are different between these two breeds, suggesting that both breeds may have been adapted to different environmental conditions. As this is the first report of the Sasso-RIR cross in on-farm testing, we encourage the continuation of testing Sasso and Sasso-RIR in a broader scale. Our results show that these breeds outperform the others, suggesting that they both cope with low-input conditions.

The Kuroiler is a dual-purpose breed developed in India under humid conditions to perform in low maintenance systems. Studies have shown its capacity to adapt to tropical countries and significantly outperform the local breeds on scavenging conditions (Sharma et al., 2015). In this study, this breed showed higher estimated body weights compared to the locally improved Horro. For the phenotypic prediction models, for all of the age phases, the body weight distribution was influenced by environmental variables linked to precipitation. Rainy season has been reported preferred for farmers to rear chickens, as high precipitation increases the vegetation and lowers predation (Bettridge et al., 2018). It also shows that the humid origin of the breed plays a role in the productivity of the breed.

The Koekoek breed is of South African origin, descending from a cross between three different breeds. This breed has been of popular use among rural farmers in African countries for egg and meat production. In a previous study using predictive habitat distribution models we showed that the Amhara region, and areas with colder temperatures and bigger fluctuations in annual mean temperature were the most suitable areas for the survival of the Koekoek breed (Lozano-Jaramillo et al., 2018). In accordance, the Amhara region and areas with higher temperature fluctuations are now predicted to have higher body weights for Koekoek males and females. Here we show, in contrast to previous results, that precipitation during the wettest and warmer periods have an influence on the predicted distribution of the breed for body weight. Even though both predictive models show temperature fluctuation to be an important variable, precipitation seems to also play a role in shaping the variation of the phenotype. Dissimilarities in the important variables that shape the predicted distribution of the breed based on presence versus the phenotype, suggests that different processes can determine the range of the Koekoeks' breed performance distribution. We believe that the use of quantitative data in phenotypic distribution models gives a more accurate result compared to the presence only models (distribution models), as the former takes into account more detailed information about how the environment influences the life history traits of the breed.

Here we show that there is variation in the body weights estimated and predicted for each of the breeds for all the growth phases. In other words, that the breeds respond differently when exposed to the same environments. Correlations between predicted and estimated body weights are positive, but for some breeds and regions they are higher than others. These correlations were calculated using the positions where both a prediction and LSmeans estimate were available. Therefore the correlation of predictions to performance outside the locations of the households in the current dataset may be lower. For some breeds the correlations are similar between regions, however for others the values greatly differ, highlighting the variation in response within breeds to different agro-ecologies.

Variation in productivity among breeds can be attributed to the breeds' origin, which can have an effect on the breeds intrinsic response to different environmental conditions. A recent study characterized the genetic diversity and identified genomic regions that presented adaptive advantages of chickens in Uganda and Rwanda (Fleming et al., 2016). Here the authors showed that the Kuroiler breed has a gene which does not occur in different African ecotypes. They indicate that this gene is associated with homeostatic regulatory functions such as response to hypoxia, cold, and starvation. They suggest that the presence of this gene may be representing the result of the selective pressures for stress tolerance during their development in India (Fleming et al., 2016).

Our results highlight the importance of taking environmental variables into account for different applications. Genotype by environment interaction (G x E), defined as the change in phenotypic performance of genotypes, relative to one another, when measured in different environments (Falconer, 1952; Falconer and Mackay, 1996) can be acknowledged using phenotypic distribution models. How sensitive a breed is to a particular environment has been of interest for many years in animal breeding. Accounting for which environmental variables have an influence on breed performance can help in predicting the occurrence of G x E interaction, and by establishing which areas within a country can be suitable for breed introduction. This knowledge can help breeding companies when choosing in which ecology they should introduce a breed, or it may help a farmer decide which breed to use in its own agro-ecology. Integrating habitat distribution models and phenotypic distribution models into animal breeding can provide a richer and more specific understanding of the environment's role in explaining the variation in productivity of a breed in different circumstances.

4.5 Conclusions

Agro-ecological diversity can be a challenge when developing breeding programs, particularly in tropical countries (Bettridge et al., 2018). The use of gradient boosting GAMs is novel approach that can be applied to this problem. They can be used either by breeding companies or by farmers to include environmental information when deciding where to introduce a breed, or which breed is better suited for which environment. We encourage the use of these models in livestock studies to capture the breeds' phenotypic variation as a response to different agro-ecological conditions. The approaches outlined in this paper can also be applied to different livestock breeds and locations and could also be used to predict impacts of different climate change scenarios on change of productivity.

4.6 Acknowledgements

The authors thank the Koepon Foundation for providing a scholarship to the first author.

4.6 Supplementary information

Table S4.1 Range of values for each of the 21 environmental variables at the locations where the breeds were tested.

Breed	Annual mean temperature (°C)	Mean temperature of warmest quarter (°C)	Mean temperature of coldest quarter (°C)	Temperature seasonality (°C)	Max temperature of warmest month (°C)	Min temperature of coldest month (°C)	Temperature annual range (°C)	Mean temperature of wettest quarter (°C)	Mean temperature of driest quarter (°C)	Mean diurnal range (°C)	Isothermality (°C)
Horro	17.95	19.39	16.62	107.21	27.43	8.30	19.13	17.60	17.07	143.42	74.60
Koekoek	18.43	20.05	17.01	119.13	28.22	8.35	19.87	18.20	17.36	147.31	73.84
Kuroiler	18.48	20.06	17.16	115.43	28.33	8.68	19.65	17.98	17.74	146.12	74.04
Sasso	18.10	19.56	16.90	107.00	27.96	8.27	19.68	17.56	17.53	146.60	74.13
Sasso-RIR	18.16	19.72	16.85	113.46	27.93	8.42	19.51	17.63	17.46	144.70	73.89

Breed	Annual precipitation (mm)	Precipitation of wettest month (mm)	Precipitation of driest month (mm)	Precipitation seasonality (mm)	Precipitation of wettest quarter (mm)	Precipitation of driest quarter (mm)	Precipitation of warmest quarter (mm)	Precipitation of coldest quarter (mm)	Total cultivated land (%)	Elevation (masl)
Horro	1075.90	210.60	12.72	76.89	536.07	60.11	245.33	217.72	32.99	2093.96
Koekoek	1023.79	227.40	11.82	88.08	562.79	52.14	222.28	161.73	33.19	2040.64
Kuroiler	1055.83	235.91	11.74	89.27	588.18	52.42	198.64	311.41	32.66	2030.03
Sasso	1273.93	278.47	13.98	84.79	702.48	60.92	240.19	334.31	31.17	2049.77
Sasso-RIR	1096.39	239.77	12.43	86.47	601.29	54.84	206.68	290.08	32.57	2076.52

Table S4.2 Range of values for each of the 21 environmental variables at the locations where the breeds were tested.

Breed	Annual mean temperature (°C)	Mean temperature of warmest quarter (°C)	Mean temperature of coldest quarter (°C)	Temperature seasonality (°C)	Max temperature of warmest month (°C)	Min temperature of coldest month (°C)	Temperature annual range (°C)	Mean temperature of wettest quarter (°C)	Mean temperature of driest quarter (°C)	Mean diurnal range (°C)	Isothermality (°C)
Horro	13-22.2	13.9-24.2	11.8-20.8	63.6-151.6	21-32.8	5.4-11.4	14.3-21.7	12.1-22.2	13.3-20.8	10.4-16.5	7.1-8
Koekoek	13-23.4	13.9-26	11.9-21.5	54.5-190.6	21-35.1	4-11.9	14.3-23.4	12.2-23.3	12.9-21.8	10.5-16.6	6.4-8
Kuroiler	13-23.4	13.9-26	11.9-21.5	70.1-188.7	21.1-35.1	4-11.9	14.423.4	12.2-23.3	12.9-21.8	10.6-16.6	6.5-7.9
Sasso	14.5-21.6	15.8-23.8	12.9-19.8	57.4-188.7	24.1-33	4-11.6	17.6-23.2	14-22.3	12.9-20.8	13.5-16	6.5-7.9
Sasso-RIR	13-23.4	13.9-26	11.8-21.5	67.6-188.7	21-35.1	4.8-11.9	14.3-23.5	12.1-23.3	13.5-21.8	10.5-16.6	6.5-7.9

Breed	Annual precipitation (mm)	Precipitation of wettest month (mm)	Precipitation of driest month (mm)	Precipitation seasonality (mm)	Precipitation of wettest quarter (mm)	Precipitation of driest quarter (mm)	Precipitation of warmest quarter (mm)	Precipitation of coldest quarter (mm)	Total cultivated land (%)	Elevation (masl)
Horro	591-1573	152-301	5-29	39-138	359-768	18-154	109-553	18-757	9.7-43.6	1667-2779
Koekoek	595-1867	146-419	2-29	39-144	359-1064	8-153	79-571	8-757	8.1-43.1	1342-2765
Kuroiler	591-1860	146-417	1-24	39-138	359-1062	4-153	65-571	14-754	11.9-43.1	1342-2762
Sasso	591-1985	150-433	2-28	48-144	359-1098	8-125	102-571	8-756	3.2-43.1	1476-2597
Sasso-RIR	591-1977	149-431	1-26	39-138	357-1093	4-153	65-340	15-757	12.8-43.1	1342-2769

Table S4.3 Correlations between the predicted values and the LSmeans for each breed in each region in Ethiopia.

Male growing period				Female growing period				Female adult period			
Breed	**Region**	**Householdsize**	**Correlation (S.E)**	**Breed**	**Region**	**Householdsize**	**Correlation (S.E)**	**Breed**	**Region**	**Householdsize**	**Correlation (S.E)**
Horro	Addis Ababa	38	0.57 (0.070)	Horro	Addis Ababa	55	0.48 (0.070)	Horro	Addis Ababa	NA	NA
	Amhara	NA	NA		Amhara	NA	NA		Amhara	NA	NA
	Oromia	NA	NA		Oromia	28	0.45 (0.103)		Oromia	32	0.65 (0.061)
	SNNPR	53	0.46 (0.074)		SNNPR	47	0.78 (0.032)		SNNPR	NA	NA
	Tigray	NA	NA		Tigray	NA	NA		Tigray	NA	NA
Koekoek	Addis Ababa	98	0.66 (0.034)	Koekoek	Addis Ababa	94	0.65 (0.036)	Koekoek	Addis Ababa	130	0.79 (0.007)
	Amhara	69	0.56 (0.053)		Amhara	71	0.55 (0.053)		Amhara	77	0.62 (0.043)
	Oromia	40	0.57 (0.068)		Oromia	44	0.47 (0.079)		Oromia	40	0.18 (0.129)
	SNNPR	101	0.81 (0.019)		SNNPR	NA	NA		SNNPR	NA	NA
	Tigray	118	0.65 (0.032)		Tigray	140	0.82 (0.015)		Tigray	140	0.85 (0.012)
Kuroiler	Addis Ababa	NA	NA	Kuroiler	Addis Ababa	NA	NA	Kuroiler	Addis Ababa	NA	NA
	Amhara	89	0.58 (0.044)		Amhara	97	0.85 (0.015)		Amhara	97	0.86 (0.014)
	Oromia	NA	NA		Oromia	NA	NA		Oromia	NA	NA
	SNNPR	98	0.73 (0.027)		SNNPR	94	0.68 (0.033)		SNNPR	94	0.78 (0.022)
	Tigray	58	0.56 (0.057)		Tigray	59	0.42 (0.075)		Tigray	59	0.34 (0.085)
Sasso	Addis Ababa	NA	NA	Sasso	Addis Ababa	NA	NA	Sasso	Addis Ababa	NA	NA
	Amhara	142	0.88 (0.010)		Amhara	155	0.8 (0.016)		Amhara	144	0.83 (0.014)
	Oromia	NA	NA		Oromia	NA	NA		Oromia	NA	NA
	SNNPR	91	0.79 (0.022)		SNNPR	124	0.81 (0.017)		SNNPR	132	0.87 (0.011)
	Tigray	NA	NA		Tigray	NA	NA		Tigray	NA	NA
Sasso-RIR	Addis Ababa	28	0.48 (0.098)	Sasso-RIR	Addis Ababa	NA	NA	Sasso-RIR	Addis Ababa	NA	NA
	Amhara	103	0.61		Amhara	76	0.6 (0.045)		Amhara	63	0.75 (0.031)
	Oromia	NA	NA		Oromia	NA	NA		Oromia	NA	NA
	SNNPR	120	0.79 (0.019)		SNNPR	103	0.78 (0.021)		SNNPR	100	0.89 (0.011)
	Tigray	64	0.29 (0.088)		Tigray	53	0.38 (0.085)		Tigray	51	0.52 (0.067)

NA= Not available or sample size less than 25.

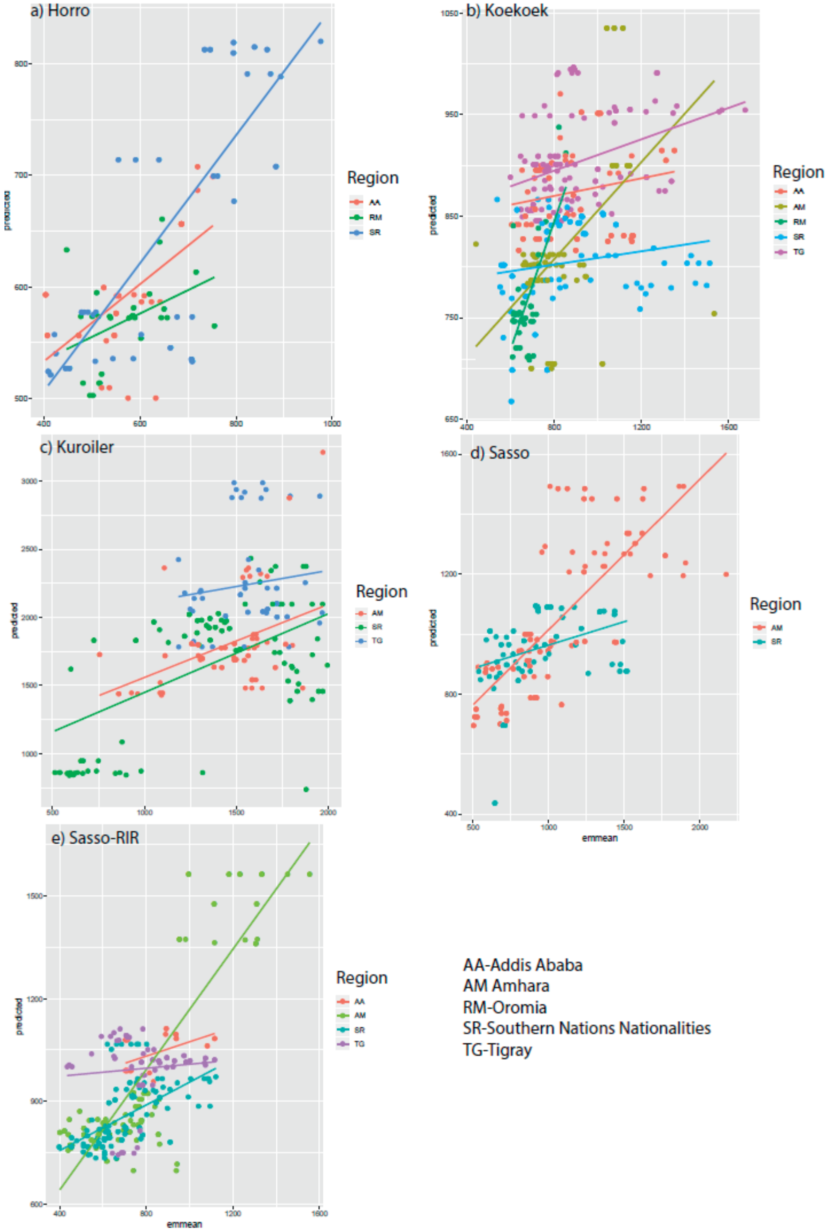

Figure S4.1 Correlations between the predicted values and the LSmeans for each breed in each region in Ethiopia for the male growing period.

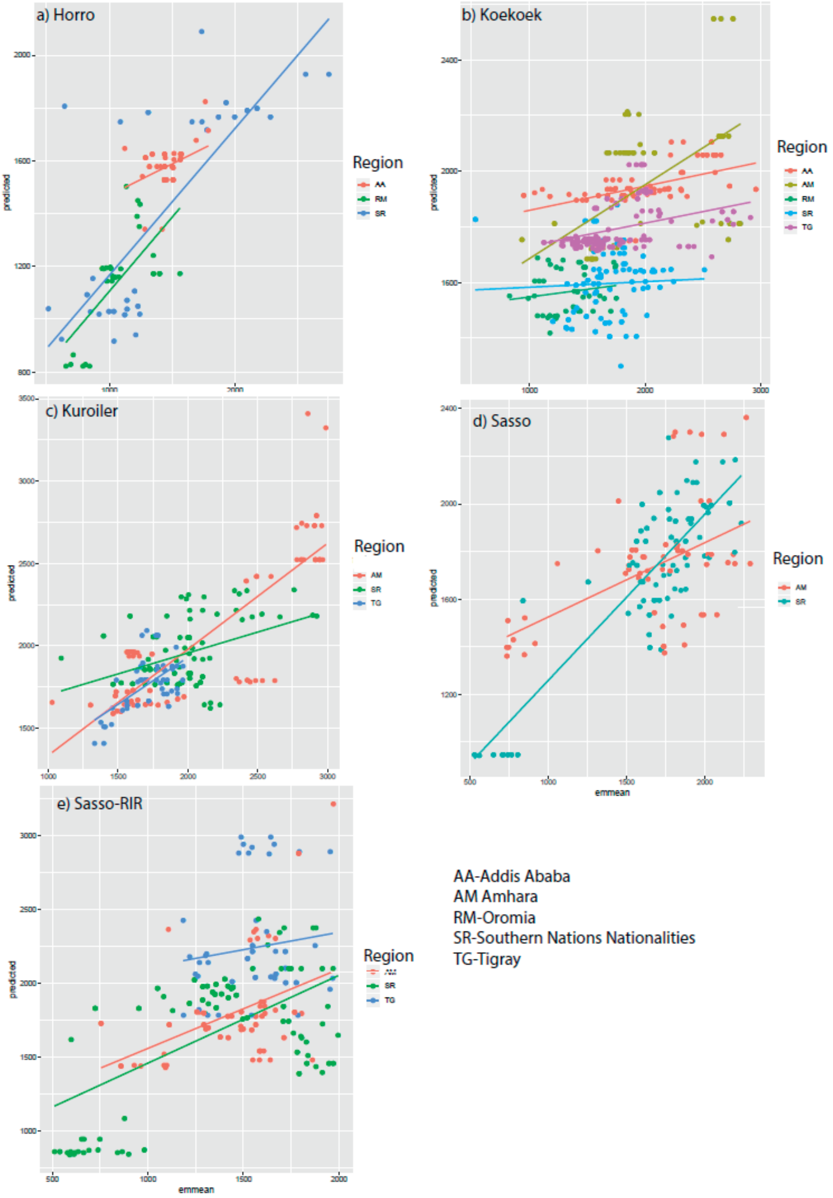

Figure S4.2 Correlations between the predicted values and the LSmeans for each breed in each region in Ethiopia for the female growing period.

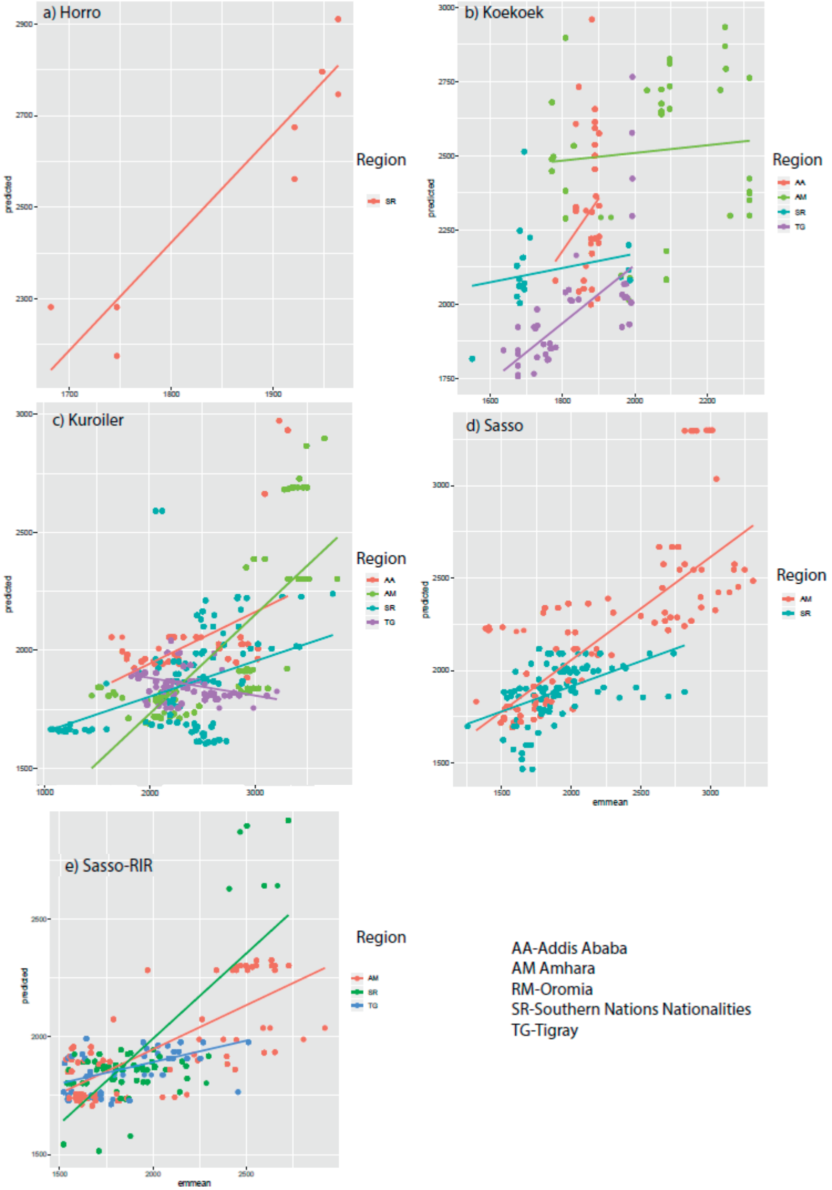

Figure S4.3 Correlations between the predicted values and the LSmeans for each breed in each region in Ethiopia for the female adult period.

References

Aman, G. et al. 2017. Management Practices and Productive Performances of Sasso Chickens Breed under Village Production System in SNNPR, Ethiopia Journal of Biology, Agriculture and Healthcare Vol 7.

Bekele, F., T. Adnoy, H. Gjoen, J. Kathle, and G. Abebe. 2010. Production performance of dual purpose crosses of two indigenous with two exotic chicken breeds in sub-tropical environment. International Journal of Poultry Science 9: 702-710.

Bekele, F., H. M. Gjøen, J. Kathle, T. Ådnøy, and G. Abebe. 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.

Bettridge, J. M. et al. 2018. The role of local adaptation in sustainable production of village chickens. Nature Sustainability 1: 574.

Bonett, D. G., and T. A. Wright. 2000. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. Psychometrika 65: 23-28.

Dana, N., T. Dessie, L. H. van der Waaij, and J. A. van Arendonk. 2010a. Morphological features of indigenous chicken populations of Ethiopia. Animal Genetic Resources 46: 11-23.

Dana, N., L. H. van der Waaij, T. Dessie, and J. A. M. van Arendonk. 2010b. 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.

Dessie, T. 2003. Phenotypic and genetic characterization of local chicken ecotypes in Ethiopia. Dessertation for Award of PhD degree at Humboldt-University, Berlin, Germany.

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

Falconer, D. S. 1952. The problem of environment and selection. The American Naturalist 86: 293-298.

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

Fleming, D. et al. 2016. Genomic analysis of Ugandan and Rwandan chicken ecotypes using a 600 k genotyping array. BMC genomics 17: 407.

Grobbelaar, J., B. Sutherland, and N. Molalakgotla. 2010. Egg production potentials of certain indigenous chicken breeds from South Africa. Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales 46: 25-32.

Habte, T. et al. 2017. Guide to chicken health and management in Ethiopia. ILRI Manual 25.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hothorn, T. et al. 2018. Mboost. R package version 2.9-1.

Lenth, R. 2019. Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.3.

Lozano-Jaramillo, M., J. Bastiaansen, T. Dessie, and H. Komen. 2018. Use of geographic information system tools to predict animal breed suitability for different agro-ecological zones. animal: 1-8.

Maloney, K. O., M. Schmid, and D. E. Weller. 2012. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. Methods in Ecology and Evolution 3: 116-128.

Mayr, A., B. Hofner, and M. Schmid. 2012. The importance of knowing when to stop. Methods of Information in Medicine 51: 178-186.

Mwacharo, J. M., G. Bjørnstad, J. L. Han, and O. Hanotte. 2013. The History of African Village Chickens: an Archaeological and Molecular Perspective. African Archaeological Review 30: 97-114.

Pym, R. 2010a. Poultry feed availability and nutrition in developing countries Food and Agriculture Organization of the United Nations, Poultry Development Review.

Pym, R. 2010b. Poultry genetics and breeding in developing countries Food and Agriculture Organization of the United Nations, Poultry Development Review.

R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved on 8 December 2016 from http://www.R-project.org.

RStudio Team. 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, USA. Retrieved on 30 June 2015 from http://www.rstudio.com/.

SASSO Breeding Company. 2018. Retrieved on November 5 from http://www.sasso.fr/home.html.

Sharma, J. et al. 2015. Higher weight gain by Kuroiler chickens than indigenous chickens raised under scavenging conditions by rural households in Uganda. Livestock Research for Rural Development 27.

Smith, A. B., J. Alsdurf, M. Knapp, S. G. Baer, and L. C. Johnson. 2017. Phenotypic distribution models corroborate species distribution models: A shift in the role and prevalence of a dominant prairie grass in response to climate change. Global Change Biology 23: 4365-4375.

Van Marle-Köster, E., and N. Casey. 2001. Phenotypic characterisation of native chicken lines in South Africa. Animal Genetic Resources Information 29: 71-78.

Wondmeneh, E. 2015. Genetic Improvement in Indigenous Chicken of Ethiopia (PhD Thesis). Wageningen Univeristy. Wageningen, The Netherlands.

Chapter 5

Predicting breed performance across countries

Maria Lozano-Jaramillo[1], Hans Komen[1], Tadelle Dessie[2], John WM Bastiaansen[1]

[1]Wageningen University & Research Animal Breeding and Genomics, P. O. Box 338, 6700 AH Wageningen, The Netherlands; [2]International Livestock Research Institute, P. O. Box 5689 Addis Ababa, Ethiopia

Submitted

Abstract

Indigenous livestock are presumed to be locally adapted. However, in tropical Africa, these local breeds do not meet the food demand required as a consequence of the rapid urbanization. Attempts to increase productivity in these areas has mainly focused on the introduction of highly productive exotic breeds. Most of these attempts have not been successful due to the non-adaptability of introduced breeds to the local conditions, underlining the fact that there is no a priori knowledge on how sensitive these breeds are to these tropical environments. Here we used phenotypic distribution models to evaluate how productivity traits of introduced chicken breeds respond as a function of the environment. Using data from Ethiopia, Nigeria and Tanzania we assessed if the phenotypic models for each breed in one country can be used to predict the performance of the same breed in the other two countries. We found differences in performance between breeds, and could attribute this to environmental sensitivity. The environmental variables that are responsible for shaping the phenotypic distribution patterns were evaluated. We show the accuracy of prediction models when projecting from one country to another, and that these models can be used to identify possible areas for successful breed introduction. The results highlight the importance of taking the environmental conditions into account when designing breeding programs for markets that encompass a range of widely different agro-ecological zones. We propose for these models to be applied to different livestock species to include the role of the environment in breed testing schemes.

5.1 Introduction

Global livestock production shows a dichotomy between the developed and developing countries. Tropical Africa, one of the least developed regions, has witnessed in the last decades a rapid urbanization, where the levels of indigenous livestock production do not meet the food demand (Thornton et al., 2009). In sub-Saharan Africa, the production systems are mostly village or semi-intensive systems, where indigenous livestock dominates in households. These livestock are typically kept outdoors, under low input conditions and with minimal health care (Moges et al., 2010; Mpenda et al., 2018). The performance of livestock under these production systems has been low for several years, as no improvements on the management are done, but rather an expansion of the herds and flocks (Thornton et al., 2009). A different approach to solve the low productivity with the indigenous livestock has been through the introduction and crossbreeding with exotic breeds. These attempts were typically done with little to no prior understanding of how these introduced breeds will adapt to the local environment and low input managements systems (Dana et al., 2010; Mwacharo et al., 2013), which probably explains why most of these attempts subsequently failed.

The African continent varies vastly in agro-ecologies. Breeds respond differently when exposed to different environments, even within countries and whether they are indigenous or introduced (Vercoe and Frisch, 1992; Montaldo, 2001; Mulder, 2007). These differences in environment and management by smallholders can expose breeds to different selective pressures (Siegel and Dunnington, 1997; Rauw et al., 1998), which may lead to genetic adaptation and result in environmental sensitivity (Lwelamira, 2012). Knowledge of the magnitude of these effects is important to determine if there will be a significant (negative) impact on breed performance at a given location, and to establish which alternative breed should be considered (Falconer, 1952; Mulder and Bijma, 2005). Furthermore, knowledge on how sensitive an introduced breed will be at a certain environment, can be used to design breeding programs tailored to a specific range of environments.

An Africa-wide collaboration project, the African Chicken Genetic Gains (ACGG; https://africacgg.net/), aims to provide better chickens to smallholder farmers in Africa. The performance of several chickens breeds introduced in three different African countries was tested, with the objective of developing a long-term chicken-breeding program under the low-input tropical conditions in each country. Knowing the sensitivity of these introduced breeds to specific environmental conditions can predict how they will perform in different environments. It will also allow for making predictions on how these breeds will respond to the different agro-ecologies throughout Africa.

Phenotypic distribution models are an approach used in ecology to evaluate how phenotypic traits respond as a function of the environment (Smith et al., 2017). Recently we used these models in livestock to predict the response of chicken body weight to different environmental conditions in Ethiopia (Lozano-Jaramillo et al., 2019). Whether these models are useful to successfully introduce breeds depends on their ability to predict the phenotypic response of an introduced breed in a completely independent specific environment outside the regions where breed performance data is known.

Our aim was to elucidate if the variation in performance of introduced breeds can be explained by linking environmental data and performance of the breed in independent regions or countries. Using performance data of two introduced breeds in Ethiopia, Nigeria and Tanzania, we evaluated if the phenotypic distribution models predicted the performance of the same breed from one country to the other countries. The three countries vary in agro-ecology (see Figure 5.1), and results have shown that within one country different environmental variables are significant in shaping the variation in breed performance (Lozano-Jaramillo et al., 2019). We evaluated which environmental variables are most important in shaping the phenotypic distribution of the breeds within each country. The phenotypic distribution models are expected to predict which locations are better suited for each breed to be highly productive, based on environmental data. Finally these models should be useful to design breeding programs, when choosing testing locations with different agro-ecologies.

5.2 Materials and methods

Phenotypic data from the ACGG project in Ethiopia, Nigeria and Tanzania was used (Figure 5.1). We estimated the phenotype-environment associations for body weight of male birds from two introduced chicken breeds. The prediction models were trained on observed data from a single country, or observed data from all three countries combined. The models were then used to predict body weight distribution in the other countries.

5.2.1 Breeds

Two commercial dual purpose chicken breeds, Kuroiler and Sasso, were tested by the ACGG project. The Kuroiler breed originates from India, and the Sasso breed from France. These breeds were chosen by the project for being adapted to tropical conditions. The two breeds were tested for growth performance in five regions in Ethiopia, five regions in Nigeria, and five regions in Tanzania (Figure 5.1). In Ethiopia approximately 350 households received chickens, in Nigeria, chickens were tested in 365 households, and in Tanzania in approximately 3200 households. Each of the households received approximately 25 six-week old chicks. A total of 7725 Kuroiler chicks and 7888 Sasso chicks were distributed in Ethiopia. In Nigeria, 6828 Kuroiler chicks, and 7884 Sasso chicks were distributed. In Tanzania, 9862 Kuroiler chicks, and 28833 Sasso chicks were distributed. The introduced chickens were kept by the farmers under the same management as the indigenous chickens. Distribution of chickens started in August 2016, and data collection ended in January 2018.

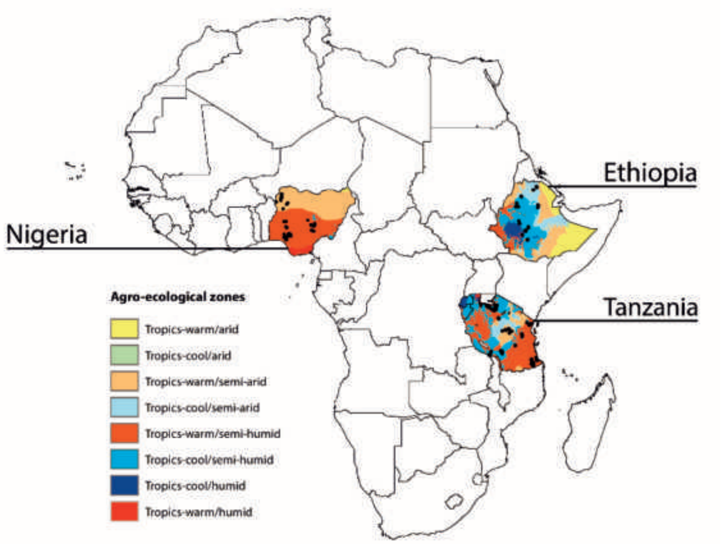

Figure 5.1 Map of the agro-ecological zones in each of the countries part of the ACGG project. Dots represent the locations were the data was collected.

5.2.2 Phenotypic data

We used live weight data from males during the growing phase (between 14-19 weeks of age). In the three countries, under traditional systems, farmers keep the birds for own consumption or they sell them at local markets when the chickens reach between 1.5-2.0 kg. Males reach this weight around week 20. Therefore, during the execution of the ACGG project, male weights were recorded until week 20.

Prior to breed distribution, each household was georeferenced. After the chickens were received, average body weight data was collected every two weeks as a group measure per household. Records with average bird weights less than 50 grams, and households with more than 30 birds reported were considered outliers and removed from the dataset. Records of birds whose age was less than 14 weeks and greater than 19 were also deleted. Data collection was not consistent, and not all weeks had records for all of the households. To solve this we standardized the body weight measures. First, for each country independently, we estimated the average week value of all measurements within a country. Then, the body weight was standardized by estimating the least squares mean (LSmeans) for each household at the averaged week. The LSmeans for all households was estimated by fitting the following linear model:

$$\boldsymbol{y} = \mathbf{H} + \mathbf{W} + \boldsymbol{e} \tag{5.1}$$

where $\boldsymbol{y}$ is the vector with the average bird weights, **H** is the household, **W** is the week of measurement, and e is the vector containing the random residual error. We assumed that within breed, all chickens had the same body weight at the time of hatching, and fitted a model with common intercept and different slope per household. For each household we calculated the least squares means using the emmeans package in R (Lenth et al., 2017).

5.2.3 Environmental data

Farmers kept the introduced chickens together with their indigenous breeds. In these village systems, chickens rely mostly on scavenging, but farmers were allowed to provide feed supplementation at their discretion. The supplemented feed was either grown by the farmers or bought externally. Under these semi-scavenging systems, food availability and chicken productivity depends on the season. To represent all this variation in environmental conditions, we used a total of 20 environmental variables at a 1km by 1km resolution that represent trends in temperature, seasonality and precipitation obtained from WorldClim (Hijmans et al., 2005). As smallholders tend to grow their own crops, and they use them to also feed their livestock, we used an additional variable representing the percentage of land used for agriculture obtained from the Harmonized World Soil Database v 1.2 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012) to represent this link between agriculture and human intervention (Dessie et al., 2003; Mwacharo et al., 2013).

5.2.4 Modelling phenotypic variation

Regression analyses are often used in ecology studies to explain the relationship between environmental variables and to build predictive models. However, ecological data can be complex showing non-linear relationships and/or spatial and temporal correlation. To address these limitations, machine learning algorithms, such as gradient boosting, have been applied to increase the prediction accuracy of standard regression models. These algorithms can address issues like overfitting and variable selection (Maloney et al., 2012).

To model the relationship between the different environmental variables and the chicken phenotypes, for each country independently, the values of the 21 environmental variables for each of the georeferenced household locations were used as predictor variables. The response variable was the LSmeans for body weight per household for each of the breeds. To increase prediction accuracy, gradient boosting, a machine learning algorithm, was applied to a generalized additive model (GAM; Maloney et al., 2012). We predicted the body weights for each country, and represented them in a heatmap for each breed. From each model the variable of highest contribution in building each of the models was obtained.

Modelling the phenotypic variation and importance variable selection were done using the package mboost in R (Hothorn et al., 2018).

5.2.5 Projecting models to another country

To evaluate if the productivity data from one country can be used to predict the productivity in other countries, we used phenotypic distribution models. The fitted model for every combination of breed and country was projected onto the other two countries. In addition, the combined data from all three countries was used to fit the model and predict the performance in each country separately. Results of the combined dataset was compared to results from models based on the individual country data.

5.2.6 Testing the accuracy of the predictions

We evaluated the accuracy of the predicted body weights by calculating the correlation between observed LSmeans and the predicted body weights by the models. For this, the predicted values from each of the projections were extracted for all the georeferenced locations of the households that received each of the two breeds. To test the accuracy, we estimated the correlation by comparing the means of the observed LSmeans and the means of the prediction data for each breed within each country.

To assess the prediction of suitable areas for breed introduction based on data from another country, the overlap between areas with high predicted body weight from data within the country, and areas with high predicted body weight from data of another country was calculated. As threshold for suitability, or high predicted body weight, we calculated for each breed-country combination the median of the observed LSmeans. If the area predicted by the other country as suitable overlaps with the areas predicted as suitable (e.g. above the threshold) by the own country data, then we consider this a successful prediction. To illustrate the area that surpasses the threshold in both predictions, we overlapped the two predicted heatmaps (Figures 5.4 and 5.5). To discover false positive predictions we overlapped the areas that were predicted above the threshold by the other country with the areas predicted below the threshold by the own country. We calculated the percentage of the whole country that is predicted to surpass the median performance threshold by both own country and other country data, while excluding the false positives. The areas of overlap would represent the locations in the country where chickens are predicted to be productive both by the own country data and also the other country data.

5.3 Results

5.3.1 Predicted body weights

The predicted body weights in Ethiopia showed that the Sasso breed had the highest mean body weight predicted (1177.9g) compared to the Kuroiler (1086.8g). The predictions for the Sasso had the widest range (412.5g-1677.5g) compared to the Kuroiler (823.2g-1500.9g). In Nigeria, the Sasso had the highest mean body weight predicted (745.8g) compared to the Kuroiler (589.4g). The Kuroiler had the widest range (358.5g-865.5g) compared to the Sasso (543.4g-814.3g). In Tanzania, the Kuroiler had the highest mean predicted body weight (1598.2g) compared to the Sasso (1312.6g), and also a widest range (1072.66g-2105.64g) compared to the Sasso (984.96g-1896.83g; Figure 5.2 and Figure 5.3; Table 5.1).

5.3.2 Importance of environmental variables

The environmental variable of strongest influence for predicting male body weight distribution was identified for each breed for each country. Patterns of male body weights for the Kuroiler breed were explained by different environmental variables across countries. In Ethiopia, the variable that had most influence in building the model was the total precipitation in the coldest three months of the year. In Nigeria, mean temperature during the driest three months of the year was most influential. In Tanzania, temperature variation between the warmest and coldest month in a year had the largest influence on the body weights distribution (Table 5.1).

Table 5.1 Median, minimum, and maximum of the observed body weights (LSmeans) for the tested households. Mean, minimum and maximum body weight predicted for the tested households across countries for different breeds. The environmental variable of importance in each of the models is shown per breed and country.

Country	**Breed**	**Median of observed body weight (g)**	**Minimum observed body weight (g)**	**Maximum observed body weight (g)**	**Mean predicted body weight (g)**	**Minimum predicted body weight (g)**	**Maximum predicted body weight (g)**	**Environmental variable of importance**
Ethiopia	Kuroiler	878.28	436.88	1678.29	1086.75	823.23	1500.94	Precipitation of coldest quarter
	Sasso	972.97	504.27	2173.21	1177.89	412.46	1677.46	Annual mean diurnal range
Nigeria	Kuroiler	510.77	205.86	972.84	589.37	358.47	865.51	Mean temperature of driest quarter
	Sasso	517.03	225.51	1054.76	745.81	543.4	814.32	Precipitation of driest quarter
Tanzania	Kuroiler	1231.04	543.76	2010.81	1598.22	1072.66	2105.64	Annual temperature range
	Sasso	1078.24	415.64	1993.74	1312.58	984.96	1896.83	Temperature seasonality

For the Sasso breed, the most influential environmental variable was also different in each country. In Ethiopia, the variable with the most influence in building the model was the mean of monthly temperature ranges. In Nigeria, precipitation during the driest three months of the year as most influential. In Tanzania, the temperature variation over the course of a year was most important to explain the predicted body weight distribution (Table 5.1).

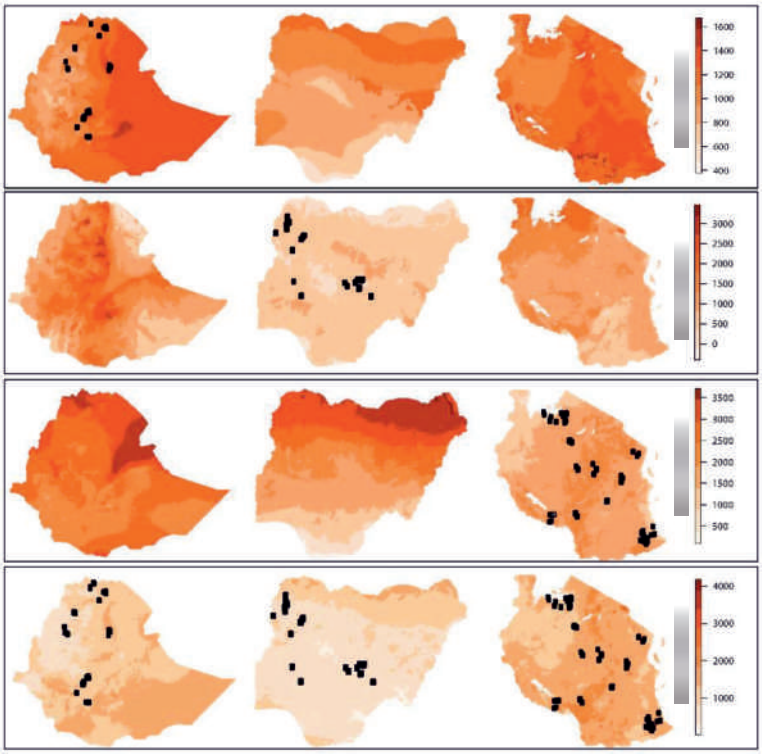

Figure 5.2 Phenotypic distribution models for body weight for the Kuroiler breed. For each row, the country with dots is used as training data to predict the body weight distribution in the three countries. In the last row, data from all three countries was used as training data.

5.3.3 Correlations and mean estimates of training data and predictions

The accuracy of the predictions was assessed by estimating the correlation between the observed LSmeans and the predictions from the projected models for each breed. Correlations between predicted and observed performance were moderate to high. The correlations were higher for Kuroiler than for Sasso, except when predicting performance in Ethiopia and Nigeria based on training data in Tanzania (Table 5.2).

Table 5.2 Correlations between the observed performance and predicted performance based on different training datasets for the Kuroiler and the Sasso, evaluated in Ethiopia, Nigeria and Tanzania.

		Training data							
		Ethiopia		Nigeria		Tanzania		3 countries	
		Kuroiler	Sasso	Kuroiler	Sasso	Kuroiler	Sasso	Kuroiler	Sasso
Evaluation	Ethiopia	0.64	0.60	0.56	0.32	0.43	0.29	0.75	0.65
	Nigeria	0.58	0.48	0.59	0.61	0.33	0.48	0.75	0.69
	Tanzania	0.48	0.46	0.57	0.54	0.58	0.56	0.64	0.66

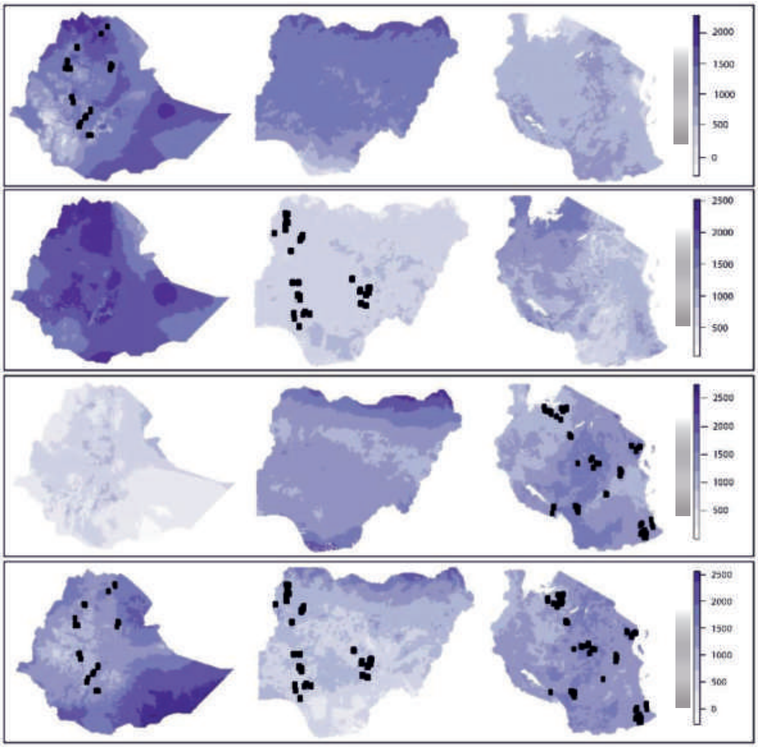

Figure 5.3 Phenotypic distribution models for body weight for the Sasso breed. For each row, the country with dots is used as training data to predict the body weight distribution in the three countries. In the last row, data from all three countries was used as training data.

To assess the prediction bias, we calculated the means of the observed LSmeans and the means of the prediction data for both breed. Performance was overpredicted for Kuroiler, except when predicting in Nigeria with data from the three countries. For the Sasso breed performance was overpredicted when predicting in Tanzania with data from Ethiopia, when predicting in Ethiopia with data from Tanzania, and when predicting in Nigeria with data from the three countries (Table 5.3).

Table 5.3 Means of LSmeans from observed performance (in the diagonal in grey), and means of the predicted performance based on different training datasets for the Kuroiler and the Sasso, evaluated in Ethiopia, Nigeria and Tanzania.

		Training data							
		Ethiopia		Nigeria		Tanzania		3 countries	
		Kuroiler	Sasso	Kuroiler	Sasso	Kuroiler	Sasso	Kuroiler	Sasso
Means of evaluation	Ethiopia	919.91	1135.91	1216.49	1801.24	1896.34	950.62	1029.32	1154.68
	Nigeria	1048.45	1191.01	556.66	609.28	1570.96	1358.36	900.07	985.63
	Tanzania	1883.6	780.07	1639.84	1124.62	1306.61	1312.42	1524.07	1234.65
		-	-	-	-	-	-	940.8	1124.93

5.3.4 Area overlap

The average value of the week for all measurement within a country was 17.1 for Ethiopia, 17.6 for Nigeria, and 16.9 for Tanzania. The LSmeans for each household averaged to the respective week showed that the median observed weight for the Kuroiler breed was 878g in Ethiopia, 511g in Nigeria, and 1231g in Tanzania. For the Sasso breed, the median was 973g in Ethiopia, 517g in Nigeria, and 1078g in Tanzania (Table 5.1). Areas within a country with a predicted body weight above the median values were considered as favorable for chicken growth. Locations where both the own and the other country predictions were above the median values, were considered areas for optimal growth. To indicate which areas were predicted as optimal for chicken growth within each country for each breed, we estimated which areas were predicted above the country’s median for each breed. To indicate which areas are predicted falsely, we overlapped the areas where the model by the other country is predicted as suitable, with the areas by the own country that are predicted as not suitable.

For Kuroiler, the projections with higher percentage of area of overlap were the predictions from the three countries together projected over Tanzania with 81.3% of overlap, and the predictions of three countries together projected over Ethiopia with 76.8% of overlap. The projections with less area of overlap were the predictions from Ethiopia projected over Tanzania with 34.3% overlap, and the predictions from Nigeria projected over Tanzania with 27.8% overlap (Table 5.4; Figure 5.4).

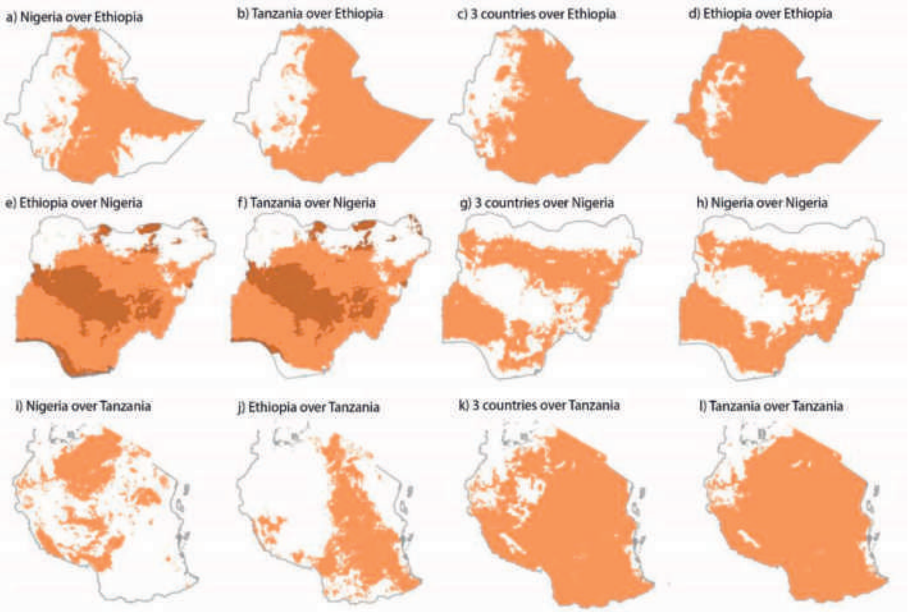

Figure 5.4 Area of overlap between the phenotypic distribution models for body weight for the Kuroiler breed from one country projected onto the other two. The first row (a-d) are the projections over Ethiopia, the middle row (e-h) are the projections over Nigeria, and the last row (i-l) are the projections over Tanzania. Darker colors denote where the model predicts areas that are not indicated as suitable by the own country (e and f).

For the Sasso breed, the projections with the highest percentage of overlap were the predictions from Tanzania projected over Nigeria with 96.2% overlap, and the predictions from Ethiopia projected over Nigeria with 94.6% overlap. The projections with the smallest area of overlap were the predictions from Ethiopia projected over Tanzania with 3.3% overlap, and the predictions from Tanzania projected over Ethiopia with 1.1% overlap (Table 5.4; Figure 5.5).

For the Kurolier, the highest within country percentage of area predicted as suitable by the model was Ethiopia with 94.1%, and the lowest was for Nigeria with 55.6%. For the Sasso, the highest within country percentage of area predicted as suitable by the model was for Nigeria with 96.2%, and the lowest was for Tanzania with 86.6% (Table 5.4).

Table 5.4 Percentage of area overlap between the prediction in one country and the prediction projected from another country per breed.

Breed	Prediction of one country projected over another	Percentage of area overlap
Kuroiler	Ethiopia over Tanzania	34.3
	Nigeria over Tanzania	27.8
	Nigeria over Ethiopia	53.3
	Tanzania over Ethiopia	70.8
	Ethiopia over Nigeria	45.9
	Tanzania over Nigeria	45.2
	Ethiopia over Ethiopia	94.1
	Nigeria over Nigeria	55.6
	Tanzania over Tanzania	86.9
	3 countries over Ethiopia	76.8
	3 countries over Nigeria	48.5
	3 countries over Tanzania	81.3
Sasso	Ethiopia over Tanzania	3.3
	Nigeria over Tanzania	31.9
	Nigeria over Ethiopia	76.5
	Tanzania over Ethiopia	1.1
	Ethiopia over Nigeria	94.6
	Tanzania over Nigeria	96.2
	Ethiopia over Ethiopia	92.7
	Nigeria over Nigeria	96.2
	Tanzania over Tanzania	89.7
	3 countries over Ethiopia	86.6
	3 countries over Nigeria	77.7
	3 countries over Tanzania	79.7

Figure 5.5 Area of overlap between the phenotypic distribution models for body weight for the Sasso breed from one country projected onto the other two. The first row (a-d) are the projections over Ethiopia, the middle row (e-h) are the projections over Nigeria, and the last row (i-l) are the projections over Tanzania.

5.4 Discussion

Environments are defined by the combination of biotic and abiotic factors of where a species lives. In nature, the combination of these characteristics define the niche for a species, and it reflects how the species copes in that environment. In scavenging systems, the combination of biotic and abiotic factors also play a role is describing how the breeds respond (i.e. its productivity). In animal breeding, when different genotypes respond differently to environmental variation is known as genotype by environment interaction (G x E). In the classical approach to estimate this response to environmental sensitivity, the environment is considered as a discrete or as a categorical unit (Calus, 2006). However, environments have continuous variation, which should be taken into account to estimate how several factors that characterize the environment have an influence on the breeds performance. Thus, when introducing one breed into another country, the breeds' environmental sensitivity should be taken into account, to know which areas can be suitable for production.

Here we integrate distribution models, as used in ecology, into animal breeding, and we found this to be effective to predict breed distribution. Gradient boosting was applied to build phenotypic distribution models to understand how environmental variables explain the variation in chicken productivity. In addition, the accuracy and bias of predicting performance based on these models was tested. Developing models that use the specific environmental variation allowed to evaluate the influence of several biotic factors on breed performance. Our results highlight the importance of taking the environmental conditions that define a location into account when testing or introducing breeds in different agro-ecologies across countries.

Within a country, the median of LSmeans for body weights of the two breeds are close in value with a maximum difference of 152.8g in Tanzania (Table 5.1). The highest body weights for both the Sasso and the Kuroiler breeds are seen in Tanzania, and the lowest in Nigeria. Within Ethiopia, heavier body weights and a wider range were predicted for the Sasso than for the Kuroiler. In Nigeria and Tanzania, the Kuroiler was predicted to be heavier than the Sasso, together with larger ranges of predicted body weights compared to the Sasso. Differences in the observed body weights across countries, and the large variation in observed body weights for both breeds, can be considered a response to the environment. Natural environments vary, and organisms can generate a different phenotypic response to this variation. When a genotype has the capacity of producing different phenotypes in different environments is known as phenotypic plasticity (Sultan, 1995; de Jong and Bijma, 2002). In animal breeding, to optimize breeding programs, phenotypic plasticity can be an important trait to select for (de Jong and Bijma, 2002). Therefore, knowing how the environment is generating these differences in phenotypic responses becomes important to have breeds that can perform in desired environmental conditions.

A study that tested Kuroiler chicken under village systems in Uganda concluded that Kuroiler succeeded compared to the other tested breeds in all the testing sites. They found that they are fast growers and perform well in scavenging systems under low management. They suggest that the Kuroiler should be a breed to keep as backyard chicken for all of Uganda, and that it should be introduced in other African countries (Sharma et al., 2015). Even though this trial was done in different regions in Uganda, the variation in the agro-ecological conditions was not taken into account to conclude that this breed could perform well in all the scavenging systems in the country. Recently we showed that the body weight of different breeds of chicken respond differently to the environmental conditions in Ethiopia, and conclude that the role of the environment in shaping the phenotypic variation should be taken into account when testing breeds in different agro-ecologies (Lozano-Jaramillo et al., 2019). Therefore, concluding that a breed can be introduced into all different areas or multiple countries should be done carefully. Scavenging systems can vary a lot due environmental variation at specific locations and whether locations are suitable for breed introduction cannot be assessed accurately without specifically taking the local environment and production system into account.

The percentage of areas that are predicted as suitable for growth within country differ between breeds (Table 5.4; Figure 5.3 and 5.4). In Ethiopia, even though most of the country is predicted as suitable for both breeds, there are different regions that are depicted as not suitable for each of the breeds. The Kurolier is not highly predicted across Nigeria (55.6%), whereas the Sasso is predicted suitable for most of the country (96.2%). The area predicted as suitable for the Kuroiler and the Sasso in Tanzania is high for both, however, the regions predicted as not suitable differ. The differences in patterns between breeds for each country, reflect that different environmental characteristics are shaping the suitability for weight distribution for each breed. Environmental suitability of areas for both breeds in each of the countries can serve as a starting point to select areas for successful breed introduction.

The best environmental predictor for the body weight patterns for each breed within and across countries was different for all breed-country combinations. The three countries share similar agro-ecological conditions, however Ethiopia is the most diverse and the least diverse is Nigeria (see Figure 5.1). These differences in agro-ecology can have an influence on the chicken feed availability, but also on the biological response to the local environment and management conditions. Management and feeding strategies also differ between and within countries. Farmers supply feed to the chickens based on their crops and/or local feeding products, which differ between countries; also as a result of differences in agro-ecologies (FAO, 2015).

Although the predictions are upward biased for most of the models, the correlations indicate that the phenotypic predictions from one country to another can distinguish suitable areas in the target country. However, there were two predictions for the Kuroiler breed where the models showed false positives; when predicting in Nigeria with data from Ethiopia and when predicting in Nigeria with data from Tanzania. For breed introduction, these areas are important to acknowledge, as there are locations where the data of the own country indicate that the breed will not perform well.

Non-adaptability of commercial breeds to different agro-ecologies has been common in breeding programs across sub-Saharan Africa (Philipsson et al., 2011). The approach presented here can be a tool to understand how different environmental conditions play a role in shaping the phenotypic variation, and can be a tool to be used for the success of breed introduction in African countries. This methodology can be expanded to other livestock, and applied as a tool to know which areas to choose for the introduction of new breeds, or to find testing sites for breeding programs.

5.5 Acknowledgements

The authors sincerely thank the Koepon Foundation for providing a scholarship to the first author. The authors would also like to thank E. Wondmeneh, S. Worku and F. Getachew and staff part of the ACGG project.

References

Calus, M. P. 2006. Estimation of genotype × environment interaction for yield, health and fertility in dairy cattle (PhD Thesis) Wageningen Univeristy, Wageningen, The Netherlands.

Dana, N., L. H. van der Waaij, T. Dessie, and J. A. M. van Arendonk. 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.

de Jong, G., and P. Bijma. 2002. Selection and phenotypic plasticity in evolutionary biology and animal breeding. Livestock Production Science 78: 195-214.

Dessie, T., C. Kijora, and K. Peters. 2003. Indigenous chicken ecotypes in Ethiopia: growth and feed utilization potentials. International Journal of Poultry Science 2: 144-152.

Falconer, D. S. 1952. The problem of environment and selection. The American Naturalist 86: 293-298.

FAO. 2015. Regional overview of food insecurity: African food security prospects brighter than ever. Accra, FAO.

FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonized World Soil Database (version 1.2), FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hothorn, T. et al. 2018. Mboost. R package version 2.9-1.

Lenth, R., J. Love, and M. Hervé. 2017. Emmeans. R package version 1.

Lozano-Jaramillo, M., S. W. Alemu, T. Dessie, H. Komen, and J. W. M. Bastiaansen. 2019. Using phenotypic distribution models to predict livestock performance In prep.

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.

Maloney, K. O., M. Schmid, and D. E. Weller. 2012. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. Methods in Ecology and Evolution 3: 116-128.

Moges, F., A. Mellesse, and T. Dessie. 2010. Assessment of village chicken production system and evaluation of the productive and reproductive performance of local chicken ecotype in Bure district, North West Ethiopia.

Montaldo, H. H. 2001. Genotype by environment interactions in livestock breeding programs: a review. Interciencia 26: 229-235.

Mpenda, F., M. Schilling, Z. Campbell, E. Mngumi, and J. Buza. 2018. The genetic diversity of local african chickens: A potential for selection of chickens resistant to viral infections. The Journal of Applied Poultry Research.

Mulder, H., and P. Bijma. 2005. Effects of genotype× environment interaction on genetic gain in breeding programs. Journal of Animal Science 83: 49-61.

Mulder, H. A. 2007. Methods to optimize livestock breeding programs with genotype by environment interaction and genetic heterogeneity of environmental variance (PhD Thesis) Wageningen University, Wageningen, The Netherlands.

Mwacharo, J. M., G. Bjørnstad, J. L. Han, and O. Hanotte. 2013. The History of African Village Chickens: an Archaeological and Molecular Perspective. African Archaeological Review 30: 97-114.

Philipsson, J., J. Rege, E. Zonabend König, and A. Okeyo Mwai. 2011. Sustainable breeding programmes for tropical low-and medium input farming systems. In: J.M. Ojango, B. Malmfors, A.M. Okeyo (eds), Animal Genetics Training Resource, version 3, 2011. International Livestock Research Institute, Nairobi, Kenya, and Swedish University of Agricultural Sciences, Uppsala, Sweden.

Rauw, W., E. Kanis, E. Noordhuizen-Stassen, and F. Grommers. 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livestock Production Science 56: 15-33.

Sharma, J. et al. 2015. Higher weight gain by Kuroiler chickens than indigenous chickens raised under scavenging conditions by rural households in Uganda. Livestock Research for Rural Development 27.

Siegel, P., and E. Dunnington. 1997. Genetic Selection Strategies-Population Genetics. Poultry science 76: 1062-1065.

Smith, A. B., J. Alsdurf, M. Knapp, S. G. Baer, and L. C. Johnson. 2017. Phenotypic distribution models corroborate species distribution models: A shift in the role and prevalence of a dominant prairie grass in response to climate change. Global Change Biology 23: 4365-4375.

Sultan, S. E. 1995. Phenotypic plasticity and plant adaptation. Acta botanica neerlandica 44: 363-383.

Thornton, P. K., J. van de Steeg, A. Notenbaert, and M. Herrero. 2009. The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. Agricultural systems 101: 113-127.

Vercoe, J., and J. Frisch. 1992. Genotype (breed) and environment interaction with particular reference to cattle in the tropics-Review. Asian-Australasian Journal of Animal Sciences 5: 401-409.

Chapter 6

General discussion

6.1 Introduction

Livestock has an important role in enhancing the food security in developing countries. They represent an invaluable source of nutrition and economic income for smallholder farmers (Hoffmann, 2010). Smallholder production systems are characterized by low producing indigenous breeds, limiting the essential food supply. As commercial breeds can be more productive than indigenous breeds, the introduction of commercial breeds into tropical Africa can benefit farmers and help in increasing the food supply. Introducing breeds improved in a centralized breeding program into a new location is a common practice in animal breeding. Nonetheless, placing new breeds into a different agro-ecology should be done cautiously. There should be a better understanding of the environmental variation that defines the biotic conditions to where the introduced breed will be exposed. In this thesis I used different approaches than what it is commonly used in animal breeding, to help understand how the environmental variation is shaping the productivity. This understanding of the effects of environmental variation allowed me to predict the performance of different genotypes in dynamic environments, which is not done in classical animal breeding schemes.

In this chapter, I will discuss how the tools developed in this thesis can be useful to understand and predict the environment’s role in shaping the phenotypic variation in livestock. I will also look at how making the link between different disciplines such as ecology and animal breeding, can enhance breeding programs. Furthermore, I recommend extending the use of these tools to improve the understanding of how climate change has an impact on the livestock sector.

6.2 Effects of dynamic environments

Natural selection is a mechanism that determines which organisms survive and pass on their genetic information to their offspring. This is generally regulated by external factors that are part of an organisms´ natural environment. Environments are dynamic, and this has an effect on populations, regardless of whether it is in breeding farms or in the wild. In animal breeding, the primary mechanism that determines which organisms survive is the consumers demand. In most production systems in developing countries however, the environment also plays an important role because of the prevalence of scavenging systems. These production systems are practiced in most rural areas, where farmers keep the livestock in their backyard allowing them to scavenge for water, household waste, insects, worms or other resources highly dependent on the local environment or season.

In the agricultural sector, climate change has generated negative impacts on the livelihoods of smallholder farmers. In sub-Saharan Africa, where economies rely mostly on agriculture, these impacts have had far-reaching consequences. Over the past decades with the increase in temperature and decrease in rainfall, the land used for agriculture has reduced significantly, generating a negative impact on food production and livelihoods of rural farmers. Environmental changes have also contributed to the decline in livestock genetic diversity, which is critical for generating food security and rural development (Hoffmann, 2010; Rojas-Downing et al., 2017). Currently there is a global focus on how livestock is contributing to climate change, but not so much on how climate change will have an impact on livestock. Strategies that work on the adaptation of breeds to future environments can help in mitigate the effect of climate change on livestock. In section 6.5 I will discuss how breeding programs can take climate change into account to design breeding programs.

6.3 The role of Geographic Information Systems (GIS) in livestock

Geographic Information Systems (GIS) are computer-based tools that help in gathering, manipulating, and visualizing data related to geographic information (i.e. geo-referenced data) usually in maps or 3D scenes. As they help in understanding spatial patterns and relationships, over the past decades, their use has been applied across a wide range of disciplines. Ecology, biodiversity conservation and geology are some examples of disciplines where its use has been fundamental. In the livestock sector, a recent interest of applying GIS tools has emerged, opening the doors to new research opportunities (Leempoel et al., 2017). Different examples include research topics related to the environmental impact of livestock, landscape and pasture management, disease control, rural development, and conservation of genetic resources (Joost et al., 2010). These differences in applications highlight the potential that these tools have in livestock research.

6.3.2 GIS as a modelling tool

GIS are powerful tools commonly used in ecology as they explore the distributions of organisms in relation to the environments where they are found. In addition, they can be used to produce models that predict the distribution of a species. A common application of GIS in wildlife is to use distribution models (Kozak et al., 2008) (also known as ecological niche models or predictive habitat distribution models (Guisan and Zimmermann, 2000)) to predict the distribution of a species across a geographic area. These models use the locations and the environmental variables at these locations, to make predictions of which other areas, based on similarities in environment, are potentially suitable for the species to live (Figure 6.1). They have served as an important tool for supporting threatened species, as they can help in delineating areas for conservation.

In this thesis I fill the gap between the distribution models commonly used in ecology and the use of GIS data in livestock. I showed how integrating distribution models to livestock research could predict current distribution of different breeds (chapter 3). This new approach provides information on the relative importance of environmental characteristics that have not been looked at in animal breeding. Now, breeding programs can use these models to tailor breeding programs to specific regions, taking into account the environmental dynamics (chapter 4 and 5). In section 6.5 I will discuss how these tools could be applied to predict breed suitability and phenotypic distribution for scenarios that take into account climate change.

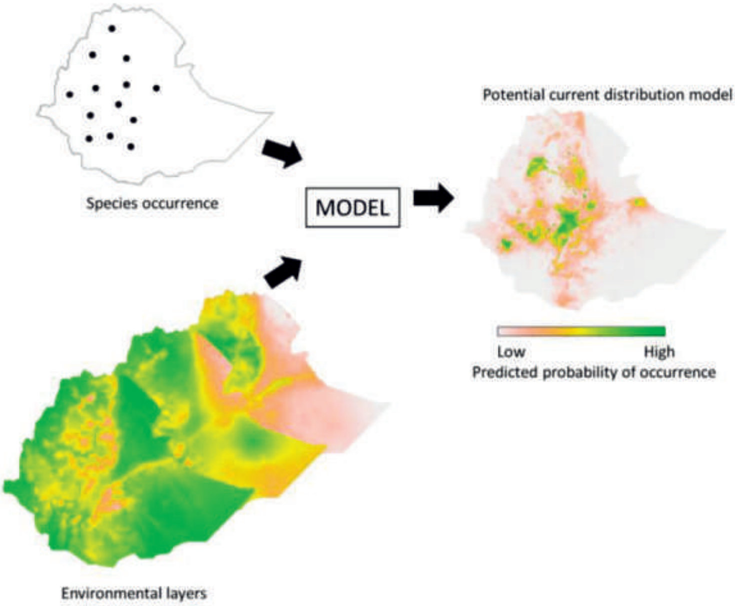

Figure 6.1 Overview of the ecological niche modelling approach. Occurrences of a hypothetical species in Ethiopia and environmental layers (in this example 3) are linked through a modelling algorithm to build the potential distribution of the species.

6.4 Role of GIS in evaluating G x E

Integrating ecology approaches to livestock research will allow making more detailed studies on genotype by environment interaction. One could potentially identify the environmental basis of breed adaptation to particular conditions and use this information to predict which areas are suitable and will be suitable under different climate change scenarios. As I have illustrated, GIS tools already serve a broad range of purposes, but in my opinion, the opportunities have not yet been thoroughly explored in livestock research. I will give some examples of the potential application on these tools in the following sections.

6.4.1 Classic G x E

The expression of a phenotype is a function of the genotype and the environment. How a phenotype is expressed, and how sensitive a breed is to a particular environment has been of interest for many years in animal breeding. Genotype by environment interaction (G x E) is defined as the change in phenotypic performance of genotypes measured in different environments (Falconer, 1952; Falconer and Mackay, 1996). Genotype can refer to a breed or individual. In this thesis I am focusing on genetic differences between breeds for environmental sensitivity. For breeding programs, G x E is important for decision making, as it indicates the environmental sensitivity of a breed. It supports the conclusion of whether a genotype is adapted for a wide range of environments, and only one breeding program is suitable, or if more than one breeding program should be implemented for different environments.

Quantitative genetic models have been, since the late 1940's, routinely implemented in animal breeding to evaluate the effects of the genetic variation in the plastic responses by animals in two or more environments (Via and Lande, 1985). In G x E studies, the environment is usually treated as discrete or categorical unit. It can be defined as different herds, different regions, qualitative differences in climates (e.g. tropical vs temperate), different feeding strategies, and different management systems (Calus, 2006). Some studies used reaction norm models to evaluate the presence of G x E as they describe the expression of a genotype as a linear function of an environmental value (Kolmodin, 2003; Lemos et al., 2015). Interactions of G x E are commonly represented as the slopes of multiple lines where the genotypes' performance are plotted against an environmental contrast. There is no G x E when the lines between genotypes are parallel, indicating that one genotype is better than the other in the compared environments. For animal breeding, the most critical scenario is when the lines between genotypes intersect. This interaction would mean that there is a re-ranking of the genotypes and that the choice for the best genotype is determined by the environment (Hayes et al., 2016; Wakchaure et al., 2016).

I will illustrate that different results can be obtained from using a classical G x E evaluation, and the models I use in this thesis. Using the data from **chapter 5**, the median of the LSmeans for body weight for the growing period (week 14-19), I evaluated if there is a re-ranking of the two breeds (Sasso and Kuroiler) between Ethiopia, Nigeria and Tanzania (Figure 6.2). In this example we see that there is no G x E between the two breeds in Ethiopia and Nigeria. The Sasso breed is performing better than the Kuroiler in the two countries. Between Nigeria and Tanzania and between Ethiopia and Tanzania there is a re-ranking of breeds. In the former comparison, the Sasso performs better in Nigeria than in Tanzania, and in the latter, the Sasso performs better in Ethiopia than in Tanzania. In both cases, the Kuroiler performs better in Tanzania. Based on this classical G x E analysis one could conclude that the Sasso should be a breed to introduce in Ethiopia and Nigeria, and the Kuroiler, a breed to introduce to Tanzania. However, as I explained before, the environment here is depicted as a categorical variable (i.e. country); the variation within the country's different environments or agro-ecologies is not taken into account. Figures 5.2 and 5.3 (**chapter 5**) show that there is variation in the body weights distributions for both breeds within each country, and I attribute this to differences in agro-ecology. Even though the overall mean body weight for each breed within each country shows differences in performance, there are some areas within Ethiopia and Nigeria that predict different patterns than what can be concluded from Figure 6.2. Hence areas in Ethiopia and Nigeria where Kuroiler can be better suited that Sasso, and areas in Tanzania where Sasso can be better suited than Kuroiler. This highlights the importance of acknowledging the environmental variation as continuous, as results from a classic G x E (e.g. using reaction norms) might lead to misleading conclusions for specific areas that are within a specific country, but that did not contribute to the dataset for the G x E.

Using GIS and ecological models, and integrating them to livestock, in this thesis, the environmental data was treated in a continuous manner. Accounting for what environmental variables have an influence on breed performance can help in predicting which other areas within a country, or across countries, can be suitable for breed introduction. In **chapter 5** I developed models that predict breed performance across countries. This means that these models can also be used to predict the performance of breeds in different countries. This will improve the results of breed introduction by reducing the chance for failed introductions, and it will result in saving time and money in breed testing for breeding programs by choosing the most informative environments.

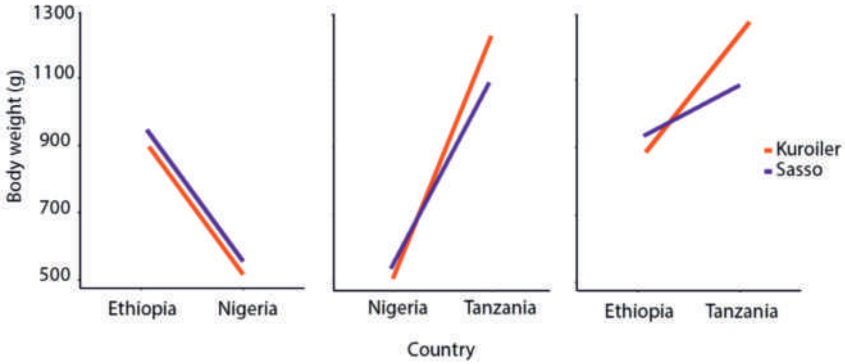

Figure 6.2 Median body weight (g) of two chicken breeds tested between 14-19 weeks of age in Ethiopia, Nigeria and Tanzania.

6.4.2. G x E using GIS

In this thesis I worked with chicken breeds, in particular with body weight. However, these models can also be used for different live history traits in chicken, such as egg number, or for different livestock species. Nowadays, breeding programs are international in the sense that breeding companies aim at having breeds that can perform in different environments. As a consequence of this globalization, acknowledging the existence of G x E becomes imperative. There are several examples of introduction programs that aim at increasing milk yield of cattle in African countries by testing individuals from nucleus farm for commercial breeds in temperate countries (Kariuki, 2017). In pig breeding, selection commonly occurs in purebred lines in temperate countries under nucleus farm conditions. For pork production, crossbred animals can be raised in temperate and tropical environments (Godinho, 2018). For both of these cases, examples are given where the magnitude of the G x E was estimated. An example where milk production from daughters from 231 common sires were evaluated in Luxemburg and Tunisia concluded that there was a re-ranking for milk yield and that milk production of daughters of the same sires depended highly on the production environment. The authors suggested that importing semen is not an effective strategy to improve milk production. They conclude that to avoid failure of breeding programs, the importing countries should have more knowledge about the environmental sensitivity of the breed to achieve the genetic progress seen in the exporting countries (Hammami et al., 2009).

In pig production, a recent study estimated the genetic correlation for carcass and growth performance traits between two different environments, temperate

(Canada, France, The Netherlands) and tropical (Brazil). The authors report that G x E appears to be present for some traits, but they conclude that the dataset, with records of over 400,000 pigs, did not yield reliable estimates for the genetic correlation between purebreds in one climate and crossbreds in the other because the pigs were too distantly related (Godinho, 2018).

As I mentioned in my thesis, I believe that a different approach that recognizes the environmental conditions within a country can be used to predict the existence and importance of G x E. For example, in cases where semen or live animals need to be exported, or in cases where performance across regions needs to be evaluated. Using the previously discussed examples, Luxemburg vs Tunisia and temperate vs tropical countries, the methodologies from this thesis should be applied as follows. Before the semen or live animals are exported from temperate to tropical or sub-tropical countries, I would first characterize the environments from both locations. For the characterization I would use different environmental layers (see Figure 6.3) such as temperature, precipitation and soil type. Then I would collect data on the locations of the nucleus commercial farms and if available, data of the current extent of the breed, hence of where the breeds are already kept by farmers. With this information, then, I would use the methodology from **chapter 3** to predict the breed suitability based on the environmental data from the current areas for the areas of desired introduction. If there are areas predicted as unsuitable in the tropical regions, then one could predict that in those areas the differences in environmental conditions would most likely lead to a G x E. Areas that are predicted as suitable in the tropical conditions could be considered as areas for potential introduction. The predictions indicate that there the existence of a G x E is not likely. These tools will lead to predicting G x E instead of having to test to find if a G x E exists.

6.5 Effect of climate change on livestock

Climate change will impose a threat to livestock production as a result of changes in feed quality and quantity, water availability, heat stress, spread of diseases, and biodiversity. Research on how climate change will have an impact on livestock production is still limited (Rojas-Downing et al., 2017). Knowledge that combines geographic, phenotypic, genotypic, and ecological attributes of livestock species is essential to prioritize the use of breeds for different environments (Joost et al., 2010).

6.5.1 Risk of livestock extinction

Biodiversity is the variability among a group of genes, organisms or ecosystems. When populations are decreasing in genetic biodiversity, they become vulnerable and at risk of extinction. One of the major drivers of biodiversity loss is climate change (UNEP, 2012; Rojas-Downing et al., 2017). Thornton et al. (2009) indicated that in the 20th century, 16% of the livestock breeds (ass, water buffalo, cattle, goats, pigs, sheep and horse) were lost. According to the FAO (2019), the world's current livestock production is based on around 40 species. As of 2018, 594 livestock breeds out of the 7,745 reported and classified as local (i.e. occurring in only one country) are now extinct. Among the extant breeds, 26% are now at risk of extinction, 67% have an unknown risk status, and 7% are not at risk. The livestock species with highest risk of breed extinction are chickens with 33%, followed by pigs with 18% and by cattle with 16%. These figures indicate that the decrease in biodiversity will have an impact on food security and rural development, and also, a decrease in genetic diversity (Thornton et al., 2009; Rojas-Downing et al., 2017). The impact of such loss in genetic diversity can have an effect on the loss of genes that are important for disease or pest resistance or to specific environmental conditions. To prevent this biodiversity loss, there should be an increase in studies that evaluate the genetic potential of breeds to adapt to future environmental conditions. With the tools developed in this thesis, researchers can benefit from choosing which areas, and which breeds to use for future scenarios. In section 6.5.3 I will show an example of how this can be done.

6.5.2 Climate change scenarios

There are different climate modelling teams around the world. To avoid the use of different models and metrics, to promote communication between groups, and ensure equal comparisons among models, the International Panel on Climate Change (IPCC) established different sets of climate change scenarios. They projected that by the end of the 21st century (2080-2100), under different greenhouse gas emissions scenarios, there will be different projected increases of global mean surface temperature. Based on different greenhouse gas concentration trajectories called Representative Concentration Pathways (RCPs), the IPCC estimates an increase of global temperature between 0.3°C-1.7°C under a scenario that assumes greenhouse gas emissions to peak between 2010-2020 (RCP2.6) and then decline. An increase of global temperature between 1.1°C-2.6°C under a scenario where the greenhouse gas emissions decline only after 2040 (RCP4.5). An increase between 1.4°C-3.1°C under a scenario where the greenhouse gas emissions decline after 2080 (RCP6.0), and an increase between 2.6°C-4.8°C under a scenario where the greenhouse gas emissions continue to rise throughout the 21st century (RCP8.5; IPCC, 2014). Depending on the region and agro-ecology, these different temperature predictions will have negative or positive effect on livestock. An increase of 2°C will produce negative impacts on feed crops, forage, and pasture in arid and semi-arid regions. However, it will have a positive impact on humid temperate regions (IPCC, 2014; Rojas-Downing et al., 2017). Therefore, different mitigation strategies should be considered depending on the region. Being able to predict the impact of these changes on livestock in arid and semi-arid regions would be of great importance for the development of policy frameworks that benefit livestock production in the future.

6.5.3 Mitigation strategies for climate change

The largest concentration of livestock productivity is found in temperate agro-ecologies (Rojas-Downing et al., 2017), whereas the largest numbers of livestock are found in the tropical and sub-tropical zones, especially in arid and semi-arid agro-ecologies (Hoffmann, 2010; Rojas-Downing et al., 2017). In these areas, environmental conditions limit the animals productivity, and increases in temperature will result in negative impacts. Heat stress is one factor that will have a negative effect on livestock production (Thornton et al., 2009; Hoffmann, 2010; Rojas-Downing et al., 2017; Hristov et al., 2018). Animals have a thermal comfort zone, which is a range of temperatures and relative humidity that are favorable for the physiological processes to be carried out. When temperature increases, surpassing the upper optimal temperature range, an animal will suffer of heat stress (Rojas-Downing et al., 2017). Acclimation is a phenotypic response to these environmental changes. However, this results in reduced feed intake, increased water intake, and different physiological functions are altered such as reproductive and production efficiency.

Finding areas where breeds will not suffer from heat stress would be beneficial to maintain optimal levels of food production, while avoiding changes on the organisms' reproduction efficiency, avoid health problems, lower the risk of mortality, and will help to avoid biodiversity loss.

Africa will be one of the continents highly affected by climate change (Niang et al., ; Figure 6.3). As the use of GIS tools has become more popular for climate change research, different environmental variables are now available to test climate change hypotheses under different temperature scenarios (e.g. RCPSs). Here I show an example on how to use GIS and the models developed in this thesis to predict effects of climate change. For this example I used the male body weight data for two chicken breeds, Kuroiler and Sasso from **chapter 5**, to predict the change in body weight as a result of climate change (Figure 6.4) under a RC6.0 climate change scenario (IPCC, 2014) projected to the year 2070.

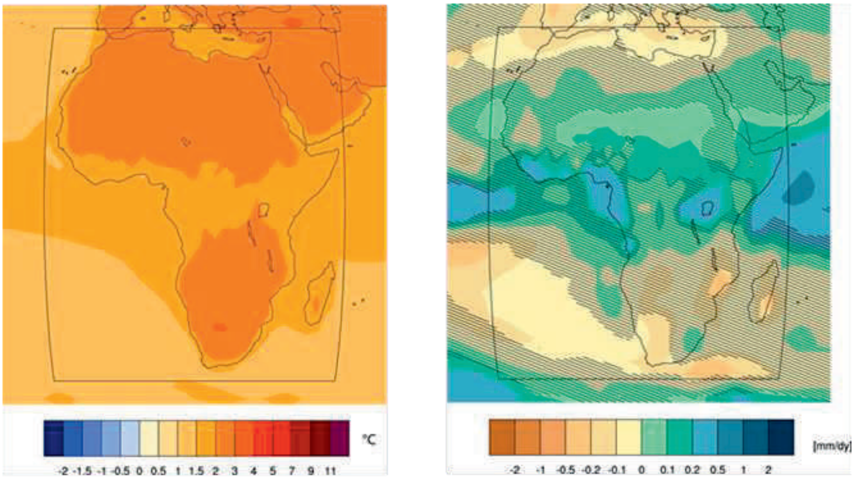

Figure 6.3 Patterns of changes in temperature (left column) and precipitation (right column) for a CMIP5 model predicted for 2050-2100 under a RCP6.0 scenario scaled by the patterns from 1986-2005. Hatching indicates not significant changes (source: CCCMA/KNMI Climate Change Alta).

In this example we see that the body weight for both breeds is predicted to decline.For the Kuroiler breed the body weight is predicted to decline by up to 30%, and for the Sasso up to 35%. For the Kuroiler, there is a predicted increase by 40% in some small areas, and for the Sasso an increase by 20% in some small areas. Birds have been reported to experience heat stress at high temperatures resulting in lower body weight gain, low feed intake, and low carcass weight (Tankson et al., 2001). Therefore, decreases in body weight as a function of high temperatures can be expected. Building these models can help in establishing mitigation strategies. More importantly, these models can also be used to design breeding programs for future climatic conditions. For example, based on the projected models for climate change (Figure 6.4), a breeding company that wants to introduce the Sasso or the Kuroiler in Ethiopia, Nigeria or Tanzania, can use this information to develop a breeding program for these predicted conditions in each country. Having information on how the body weight will change in response to the environment, commercial companies can use the tools developed in this thesis to find current areas that will resemble the future environmental conditions and use this information to select perform breed testing for heat resistant breeds.

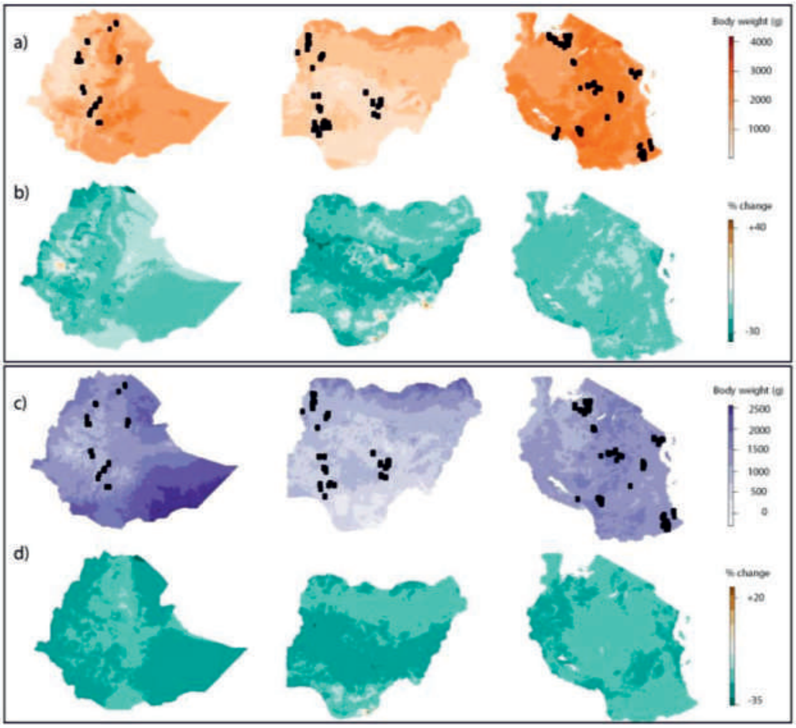

Figure 6.4 Predicted current body weight measures for the a) Kuroiler and the c) Sasso breeds for Ethiopia, Nigeria and Tanzania. Dots represent the locations where the data was collected. The predicted percent change in body weight for the b) Kuroiler and the d) Sasso was calculated as (future prediction value divided by current predicted value)*100.

6.6 Landscape genomics: opportunities in livestock

Evolutionary biology and population genetics aim at understanding the spatial distribution of adaptive genetic variation in ecosystems. Inferring these evolutionary processes is challenging, as ecosystems are complex and spatially and temporally changing. The use of molecular markers have helped in facing these challenges by making more robust evolutionary insights on local adaptation (Pariset et al., 2012). An emerging field that combines genetics with the landscape structure, landscape genomics, has offered the possibility of better understanding the patterns of the geographic distribution of the genetic variation. Landscape genomics correlates the patterns of genetic variation with geographic variables, to evaluate how geographical and environmental characteristics shape the genetic structure of populations (Pariset et al., 2012; Li et al., 2017). This field has over the past decade, started to become popular in livestock research.

Landscape genomics becomes an important tool to identify genotypes that are suited to a specific environment (Joost and Negrini, 2010). Having this knowledge can help understand which genotypes can adapt better to different environments or climate change scenarios, therefore, helping in mitigating the effects of climate change on livestock. Confounding effects between genotypes and climatic variables generates opportunities for adaptation to take place. Therefore, I believe that landscape genomics can be used to understand the associations between environments and genotypes for livestock production. For these cases, as suggested by Leempoel et al. (2017), GIS can be an important tool to relate the environmental variables to the locations where the breeds are located, and to relate the genomic information to find genes that show evidence of adaptive variation.

If a researcher has an environmental characterization of the breeds landscape (niche), and associate these variables to the genetic variation of a breed (i.e. SNPs), then they can see if there is a correlation between environmental variables and the genes which are significant for shaping different life history traits. With this information they could evaluate how traits will be spatially distributed for a breed based on environmental variability. To further develop the tools and results from this thesis I could incorporate genetic information to associate the genetic structure of the breeds to the productivity traits and the environmental variables. I could go further and finds associations between the phenotypic and the genomic information and find if this associations are also correlated with the environmental variation.

6.7 Concluding remarks

In this thesis I find that predictive distribution models are a powerful tool to predict animal breed suitability for different agro-ecologies. Therefore, integrating ecological models in livestock research is a novel approach that can have a positive impact on different decision-making strategies in animal production. Predicting the existence of G x E, or selecting testing sites for breeding programs by the generation of models that describe the breeds' sensitivity to a range of environmental conditions can be an important tool in animal breeding and livestock production. Defining testing sites can be done for current environmental conditions as was explained in the thesis (**Chapter 3-5**). Selecting locations for future climatic conditions, as explained in this chapter, can lead to the generation of mitigation and conservation plans.

Integrating different disciplines such as ecology and landscape genomics will rely on collaborations across multi-disciplinary groups. This integration can be a challenge, but it will help in understanding the role of the environment in local adaptation, and at the same time will help in enhancing food security and biodiversity conservation.

References

Calus, M. P. 2006. Estimation of genotype × environment interaction for yield, health and fertility in dairy cattle (PhD Thesis) Wageningen Univeristy, Wageningen, The Netherlands.

Falconer, D. S. 1952. The problem of environment and selection. The American Naturalist 86: 293-298.

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

FAO. 2019. The State of the World's Biodiversity for Food and Agriculture, J. Bélanger & D. Pilling (eds.). FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome. 572 pp. (http://www.fao.org/3/CA3129EN/CA3129EN.pdf)

Godinho, R. 2018. Genotype by environment interaction for feed efficiency in growing-finishing pigs in Brazil versus the Netherlands (PhD Thesis) Wageningen Univeristy, Wageningen, The Netherlands.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Hammami, H. et al. 2009. Accessing genotype by environment interaction using within-and across-country test-day random regression sire models. Journal of Animal Breeding and Genetics 126: 366-377.

Hayes, B. J., H. D. Daetwyler, and M. E. Goddard. 2016. Models for genome× environment interaction: Examples in livestock. Crop Science 56: 2251-2259.

Hoffmann, I. 2010. Climate change and the characterization, breeding and conservation of animal genetic resources. Animal genetics 41: 32-46.

Hristov, A. N. et al. 2018. Climate change effects on livestock in the Northeast US and strategies for adaptation. Climatic change 146: 33-45.

IPCC. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.

Joost, S. et al. 2010. Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. Animal Genetics 41: 47-63.

Joost, S., and R. Negrini. 2010. Early stirrings of landscape genomics: awaiting next-next generation sequencing platforms before take-off. 9251066973, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture.

Kariuki, C. M. 2017. Breeding strategies for sustainable intensification of developing smallholder dairy cattle production systems (PhD Thesis) Wageningen University, Wageningen, The Netherlands.

Kolmodin, R. 2003. Reaction norms for the study of genotype by environment interaction in animal breeding (PhD Thesis) Swedish University of Agricultural Sciences No. 437, Uppsala, Sweden.

Kozak, K. H., C. H. Graham, and J. J. Wiens. 2008. Integrating GIS-based environmental data into evolutionary biology. Trends in ecology & evolution 23: 141-148.

Leempoel, K. et al. 2017. Simple Rules for an Efficient Use of Geographic Information Systems in Molecular Ecology. Frontiers in Ecology and Evolution 5.

Lemos, M. et al. 2015. Reaction norms for the study of genotype-environment interaction for growth and indicator traits of sexual precocity in Nellore cattle. Genetics and Molecular Research 14: 7151-7162.

Li, Y. et al. 2017. Ten Years of Landscape Genomics: Challenges and Opportunities. Frontiers in Plant Science 8.

Niang, I. et al. 2014: Africa. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L.White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1199-1265.

Pariset, L., S. Joost, M. Gargani, and A. Valentini. 2012. Landscape genomics in livestock In: Analysis of Genetic Variation in Animals (ed. by M. Çaliskan). p pp.189-216. IntechOpen, Rijeka, Croatia.

Rojas-Downing, M. M., A. P. Nejadhashemi, T. Harrigan, and S. A. Woznicki. 2017. Climate change and livestock: Impacts, adaptation, and mitigation. Climate Risk Management 16: 145-163.

Tankson, J., Y. Vizzier-Thaxton, J. Thaxton, J. May, and J. Cameron. 2001. Stress and nutritional quality of broilers. Poultry Science 80: 1384-1389.

Thornton, P. K., J. van de Steeg, A. Notenbaert, and M. Herrero. 2009. The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. Agricultural systems 101: 113-127.

UNEP. 2012. (United Nations Environment Programme). Global environment outlook 5: Chapter 5. (https://wedocs.unep.org/bitstream/handle/20.500.11822/8021/GEO5_report_full_en.pdf).

Via, S., and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39: 505-522.

Wakchaure, R., S. Ganguly, and P. Praveen. 2016. Genotype X Environment Interaction in Animal Breeding: A Review. Biodivers. Conserv. Chang. Clim: 60-73.

Summary

Summary

In most of African countries, livestock production branches from an ancient tradition where nearly all rural and peri-urban families keep different indigenous breeds in scavenging systems. In sub-Saharan Africa, where these production systems are the most prominent, livestock mainly forages for resources that are highly dependent on the local environment and season. Even though these breeds are said to be adapted to the local conditions, their productivity is still low compared to commercial breeds. There have been several efforts from researchers, policy makers and livestock specialists to introduce commercial breeds to support the generation of food security and poverty alleviation. However, most of these attempts have failed because of the non-adaptability of introduced breeds to the local conditions. Typically there is no prior knowledge on the environmental sensitivity from these breeds to this new tropical environments. Throughout this thesis I use Geographic Information Systems (GIS) that describe the environment, and models used in ecology to investigate the match of animals with their environment. The aim of this thesis was to evaluate how the environment plays a role in shaping differences in breed performance across agro-ecological zones, and what implications this can have in genetic improvement of livestock.

Several animal breeding studies tested breeds in different environments to evaluate whether genotypes respond differently to changes in the environment (i.e. G x E). To estimate if there is a re-ranking in breed/genotype performance between environments, a genetic correlation is estimated. To accurately estimate this correlation, an optimal mating design should be established. Breeding programs use full-sibs or half-sibs to perform testing in different environments. Within families, common environmental effects can be present generating a covariance between siblings, and should therefore be taken into account when estimating genetic correlations. In **chapter 2**, I used stochastic simulation to find the optimal population structure to accurately estimate the genetic correlation between environments using a combination of full-sibs and half-sibs groups under different levels of common environmental effects. Simulation results showed that when there are no common environmental effects present in the population, the mating ratio that gives the lowest standard error of the genetic correlation is of one female per male with 10 offspring per sire per environment. Not accounting for common environmental effects when these are present in the population will lead to an upward bias of the genetic correlation. Increasing the number of females per male to a minimum of 5, with 10 offspring per sire per environment will alleviate

the impact of common environmental effects lowering the standard error of the genetic correlation. I suggest for studies that aim to estimate the magnitude of the G x E, to acknowledge the presence of common environmental effects and to take this into account when deciding the mating ratio.

In **chapter 3**, using GIS and habitat distribution models a methodology to predict breed suitability for different agro-ecological zones was developed. The methodology was tested on the current distribution of two introduced chicken breeds in Ethiopia. Results show that this methodology is effective in predicting breed suitability for specific environmental conditions. For both chicken breeds the model predicts suitable areas beyond their current extent, hence suggesting areas for breed introduction. The most significant variables that explain the current breed distribution were similar to the environmental conditions from which the breeds originate.

In **chapter 3**, only information on the location of the breeds was taken into account. This was extended in **chapter 4**, leading to an approach that predicts the productivity of the breeds. I present a methodology to model breed performance by using growth data from five different introduced chicken breeds in Ethiopia part of the African Chicken Genetic Gains project (ACGG; https://africacgg.net/). The suitability of these breeds was tested by predicting the response of body weight as a function of the environment in Ethiopia. Across the Ethiopian landscape, predicted body weights varied for all of the breeds. The variation in body weight was explained by different environmental variables, highlighting the importance of understanding the role of the environment in predicting breed productivity.

In **chapter 4**, breed performance was predicted within a single country. In **chapter 5** breed performance was predicted across countries. Growth data was used from two chicken breeds that were introduced in Ethiopia, Nigeria and Tanzania by the ACGG project. The aim was to assess if the data from one country could be used to predict the performance of the same breed in the other two countries. The variation found in breed performance could be attributed to each breeds' environmental sensitivity. The environmental variables responsible for shaping the variation in performance were different for each breed in each country. The accuracy of the prediction models projected from one country to the other show they can be used to identify areas for successful breed introduction.

In **chapter 6** I discussed how the tools developed in this thesis can be used in animal breeding for different approaches. I suggest for different disciplines such as landscape genomics and ecology to work together with animal breeding to understand the role that the environment plays in shaping the observed phenotypic differences. This knowledge has implications for the development of breeding programs for different agro-ecologies, taking into account the continuous environmental variation. Furthermore, I recommend the use of these tools to generate knowledge on the impact of climate change on livestock to help generate mitigation plans and policy frameworks that will help in enhancing food security and preserving the current biodiversity.

Curriculum vitae

About the author

María Lozano Jaramillo was born on the 30th of January 1985 in Bogotá, Colombia. In 2006 she obtained an associate in arts degree in biology from the Broward College in Florida, USA. In 2010 she obtained a bachelor in biology from the Universidad de los Andes in Bogotá, Colombia. In 2013 she obtained her masters degree in biological sciences from the Universidad de los Andes in Bogotá Colombia. During the bachelor and masters her focus was in evolutionary biology and ecology. Her studies focused on integrating molecular and geospatial approaches to understand how a disjunct distribution of a species of hummingbird can be originated and maintained. For her master's she focused on integrating phylogeographic, population genetics, and physiological approaches to understand the role that adaptation and natural selection could have had in resolving the challenge of living at high elevations for a species of duck in South America. On 2013 she worked as a teacher at Gimnasio la Montaña in Bogotá Colombia. On 2014 she worked at The Alexander von Humboldt Institute For Research on Biological Resources in Bogotá, Colombia as a climate change mitigation research assistant. On 2015 she started her PhD in the Animal Breeding and Genomics group at Wageningen University in The Netherlands. As part of her research, she spent 14 months at the International Livestock Research Institute (ILRI) in Addis Ababa, Ethiopia. During her PhD research she was responsible for analysing and gathering data part of the African Chicken Genetics Gain program. The aim was to understand how the environment plays a role in shaping differences in breed performance across different agro-ecological zones, and what this means for genetic improvement of chickens. The results of this research are presented in this thesis.

Publications

Lozano-Jaramillo, M., Alemu, S.W., Dessie, T., Komen, H., &. Bastiaansen, J. W. M. 2019. Using phenotypic distribution models to predict livestock performance. EAAP contribution.

Bettridge, J. M., Androniki P, Zelalem G.T, Takele T.D, **Lozano-Jaramillo, M.**, Dessie, T., Kaiser, P., Wigley, P., Hanotte, O., & Christley, R. M. 2018. The role of local adaptation in sustainable production of village chickens. Nature sustainability, 1,10: 574.

Lozano-Jaramillo, M., Bastiaansen, J. W. M., Dessie, T., & Komen, H. 2018. Use of geographic information system tools to predict animal breed suitability for different agro-ecological zones. animal, 1-8.

Lozano-Jaramillo, M., McCracken, K. G., & Cadena, C. D. 2018. Neutral and functionally important genes shed light on phylogeography and the history of high-altitude colonization in a widespread New World duck. Ecology and Evolution.

Lozano-Jaramillo, M., Bastiaansen, J., Dessie, T., & Komen, H. 2018. Understanding the effect of the environmental conditions on the suitability of a breed for different agro-ecological zones. WCGALP contribution.

Lozano-Jaramillo, M., Rico-Guevara, A,. & Cadena, C. D. 2014. Genetic differentiation, niche divergence, and the origin and maintenance of the disjunct distribution in the Blossomcrown *Anthocephala floriceps* (Aves, Trochilidae). PLoS One: 9: e108345.

Munoz-Fuentes, V., Cortazar-Chinarro, M., **Lozano-Jaramillo**, **M** & McCracken, K., 2013. Stepwise colonization of the Andes by Ruddy Ducks and the evolution of novel β-globin variants. Molecular Ecology 22: 1231-1249.

Lozano-Jaramillo, M. Variación intraespecífica y el origen y mantenimiento de la distribución disyunta del Colibrí Cabecicastaño *Anthocephala floriceps* (Trochilidae). Ornitología Colombiana 10: 75.

Book chapters

Morales-Rozo, A., Rico-Guevara, A., Moreno-Palacios, M., & **Lozano-Jaramillo, M**. (2016) *Anthocephala floriceps*. in: Rengifo L.M., Amaya-Villareal, A.M., Burbano-Girón, J. Velásquez-Tibatá, J (eds). 2016. Libro rojo de aves de Colombia, Volumen II: Ecosistemas abiertos, secos, insulares, acuáticos continentales, marinos, tierras altas del Darién y Sierra Nevada de Santa Marta y bosques húmedos del centro, norte y oriente del país. Publisher: Pontificia Universidad Javeriana. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, pp.196-198.

Rico-Guevara, A., Morales-Rozo, A., Moreno-Palacios, M., & **Lozano-Jaramillo, M**. (2016) *Anthocephala berlepschi*. in: Rengifo L.M., Amaya-Villareal, A.M., Burbano-Girón, J. Velásquez-Tibatá, J (eds). 2016. Libro rojo de aves de Colombia, Volumen II: Ecosistemas abiertos, secos, insulares, acuáticos continentales, marinos, tierras altas del Darién y Sierra Nevada de Santa Marta y bosques húmedos del centro, norte y oriente del país. Publisher: Pontificia Universidad Javeriana. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, pp.200-203.

Grants

Travel Award LEB Foundation. Wageningen, The Netherlands. (2018)
Young Scientist Scholarship. 11th World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand. (2018)

Travel Award. Meeting of the American Ornithologists' Union, Cooper Ornithological Society, and Society of Canadian Ornithologists. Chicago, Illinois. (2013)

Colombia Biodiversa. Fundación Alejandro Ángel Escobar, Bogotá, Colombia. (2012)

Chapman Grant. American Museum of Natural History, New York, USA. (2012)

Research project grant for graduate studies in biology. Universidad de los Andes, Bogotá, Colombia. (2011)

Training and supervision plan

EDUCATION AND TRAINING

A. The Basic Package

WIAS Introduction Day	2018
Course on philosophy of science and/or ethics	2016

B. Disciplinary Competences

Writing Research Proposal	2015
Biological Consequences of Selection	2015
Advanced course on “Design and implementation of breeding programs for smallholder poultry farmers”	2015
Fundamentals of Data Analysis and Mixed Linear Models in Animal Breeding	2016
Genetic Improvement of Livestock	2015
In-depth Course Genotype by environment interaction, uniformity and stability	2015
Statistics for the Life Sciences	2016
Getting started in ASReml	2017
Statistical Analysis of On-Station and On-Farm longitudinal Chicken Performance Data Using R Sofware	2017

C. Professional Competences

Writing Grant Proposals	2016
Brain Training	2016

D. Presentation Skills

World Congress on Genetics Applied to Livestock Production (WCGALP), New Zealand-oral	2018
WIAS Science Day, The Netherlands-oral	2019
European Association of Animal Production (EAAP)-oral	2019

E. Teaching competences

Teaching Assistant for the Animal Breeding and Genetics course	2016
Supervision minor MSc thesis	2019

Acknowledgements

Acknowledgements

Four years have gone by fast. A PhD is never done alone. There is a big team of people backstage that should be acknowledged. One of my biggest motivations in life is my grandmother. I owe everything I am to that wonderful, amazing, inspiring woman. I cannot start thanking everyone without giving her special thanks. Abue, eres la mujer más incredible del mundo. Te admiro infinito. Gracias por siempre estar ahí, por apoyarme en todo sin medida, por tu generosidad y por cuidarme todos estos años. Esta tesis no hubiera sido posible sin ti. Te amo con todo mi corazón. A mi mamá, donde quiera que estés, gracias por tus enseñanzas. Este logro sigue siendo parte de tu apoyo y amor incondicional. Te pienso cada día. Te amo. I am also indebted to the rest of my beautiful family. Even though we are scattered all over, we manage to stay together. Thank you all for being always present. A mis tías, Tata y Gogas, gracias por su apoyo incondicional. Las quiero mucho. A mi tío, Mari, Manu y Antonia, son una hermosa familia. Los quiero muchísimo. Gracias por estar pendientes, por las charlas, las llamadas, las visitas aquí en Holanda o el hospedaje en Londres. Juanis, gracias por apoyarme en todos los momentos, buenos y malos. Gracias por las charlas, por visitarme, alojarme y por nunca dejarme sola. Andreas, gracias por cuidar a Juanis. Mis cabritas, siempre presentes. Gracias por todo el apoyo en todo y por estar siempre pendientes. Linis, Tuna, Ire, Paulis, Nanis, Silvi y Caro, gracias por tanto amor y tanta generosidad.

The main reason I came to Wageningen is because I had an interview with Elise and Liesbeth. Thank you Elise and Liesbeth for believing in me and opening these doors. Johan, thanks for trusting Elise and Liesbeth, and for giving me this opportunity. You were always involved in my project, one way or another, thanks for this support. In these four years, I jumped between Ethiopia and the Netherlands. It was not always easy leaving things behind in both countries, but it was always nice to see familiar faces in both places. John and Hans were always around in this process from Wageningen, and Tadelle from Addis. Hans, thanks for being always involved in this endeavor. It is very hard to summarize four years of working together, but I can say it was great to have you as a promotor. You were always actively involved, in charge of the project, and looked after me. It was nice travelling to Ethiopia together, and good to share valuable moments in the Netherlands. We had a great four years together; you as yellow, with your door always open, me also yellow, always talking. All this “yellowness” was indeed very ”gezellig”. But as nice as it is to work with yellow people, it was great to add some blue to this equation. John, thank you for these four years of learning, teaching and patience. Thanks for always having your door open, and for our talks, which always started with half an hour of catching up. I learn (still do) a lot from you. It has been amazing to work with together. I admire your passion for work, and to learn new things, and how you love and take care of your family. Thanks for all of your teachings. Thanks also to the Koepon Foundation for funding my PhD, and to Linda and Gerard for your interest in my role in the project. Lis and Adita, you two had a very important role in these four years. I really appreciate all your help, which was always done with a smile. Thanks for caring about my things, and for caring about me; for looking after me when I was in the Netherlands or in Ethiopia. Rosilde, thanks for managing the finances. Lucia, thanks for trusting and believing in me. Richard, it has been great to work with you these past few months.

Many people come and go in Wageningen. Some stay, some come back, and with some I managed to stay in touch. Somehow, they all played a role in these four years. The first person from ABG that I met was Andre, my buddy. Thanks for introducing me to everyone and to the group dynamics. My first corridor mates, Jovana and Sandrine, thanks for allowing only happy people there. Andre, Vivi, Marconha, thanks for introducing the “caipirinha de abacaxi” to Yvonne, and for all the nice parties, dinners and times together. Mathieu and Laura thanks for being there. Cori, Sonia, Cori, and Gareth, thanks for the great times we had together. It was great sharing an office together, even though we did not let Ole work next door with our loud voices (mostly coming from Zih-hua). You managed to finish your thesis in time, and to teach me all the cursing in French and Afrikaans (we even kept a score board of this). Thanks ladies, including Crystal, for all the support always, for being there even from a distance. Gracias Cori por todo el apoyo, que también hemos logrado mantener desde la distacia. Mathijsito, thanks for letting us crash at your place, for the dinners, but mostly for letting me play the violin when Cori was in town. After this office, I was part of a more tropical environment with the Maluma babies, Clau and Rodrigo. This was a great office, and I mean this not only because we listened to lots of reggaeton (some Ed Sheeran and Kylie Minogue too), and spoke Spanish (and Portuñol), but also because we became a little family (las feijoadas y pão de queijo también tienen que ver). Clau, gracias por estar siempre pendiente de mis cosas, por enseñarme de perseverancia. Fue muy especial haber compartido tantos viajes contigo. Es muy lindo verte con Pierre y Guille. Son un ejemplo enorme para mi. Los quiero mucho. Hermanito, eres un sonso que amo! Gracias por ser tan especial, por aguantarte todas mis historias (sin juzgarme) y por confiar en mi.

After this tropical period came the Asian office (Chiarina counts as Asian too, sometimes). I came into this office already after you were all together for some

time, but still I was invited to be part of your cooking evenings with board games! I loved these evenings. Even though, I must confess, I might not be able to eat couscous again ;). Thank you Harmenito for being attentive and sweet always. Thanks for the board games, for caring about me, for learning with me to say “hello” or “good morming” in all the languages that we have in the group, and for teaching me to be a “mierenneuker” with PowerPoint. Remember, you are not a colleague, but a friend. Zhoucita, thanks for always having a smile, for running around like an ostrich, for the endless amount of medicine, food, and random things you have in your closet, and thanks for always being helpful and positive. Vinicius, it was good to share this last stretch together (except for when you panicked because of my explanations). Langquincito, thanks for all the interesting political discussions, but mostly for the meditations tips, and cooking shows, “tengo hambre does really match with you ;). Thanks for being also sweet always, for being always “tranquilo”, even when playing foosball. Thanks Chiarina for being special, caring, and for liking me talking in the office (you do like it a bit, right?). Thanks for your honesty and for being direct. I admire your perseverance and ambition. Thanks to all this office for the political, societal, linguistic, and random discussions, which led to our biology class board.

Outside of the “office sharing space”, there are some people that I did not share an office with, but for some reason we became close. Maybe it was because of the parties, partycom events, pub quizzes, or just because of the “gezelligheid”. Seyda, was really nice to work with you, even though it was for a short time, I learned a lot from you. As ABG is so big, I’m probably forgetting people. Thanks to everyone in the group for the great times. Jeremie, Kasper, Henk, Mario, Piter, Martien, Benan, Yvette, Anouk, Aniek, Samuel, Florecita, Martincito, Biaty, Ibrahim, Haibo, Michael, Malou, thanks for all the nice talks and biertjes. Marieke, thanks for always being sweet, for always having a smile, and for letting me join you in class :). Henricito and Laura, apart from the drinks, it was nice to try to make buñuelos. Jani, it is really nice that you joined the group. It has been so much fun to spend time together. Limcita, thanks for all the laughs and surprise visits in our office. It was great to spend time during the EAAP with you and our little cloud. Renzito, my postman, thanks for all the fun, and for being so sweet. Shuwen, my Chinese cousin, you are one funny cousin. Thanks for always being there, hungry or sleepy, you were always there. It was great to be a last year PhD candidate with you. I say this, not only because we were working on Saturdays together, but also because it was amazing to help and support each other. Sanne, thanks for caring about me and about my things. Thanks for driving me around when I was injured, and for helping me always. Yvoncita, you are amazing. Thanks for all your help with everything. You are an amazing friend, great cripple taxi driver, and a great co-author. It´s has been amazing sharing all these different phases together. I´m very happy about upgrading “our” B&B ;). Pascalito, it was great to spend these four years together. I mean this not only because of all the parties, dancing, and travels, but because you became a close friend. Thanks for caring about me, and for listening to all my “dramaria” stories. I am sure we will have plenty of parties and trips yet to come.

Outside of the group, there are also people that are always present, and that were part of these years. Chino, gracias por ser tan buen amigo, por ayudarme siempre que estaba sin donde dormir, por invitarme a los asados peruanos, a almorzar, al cafecito, de rumba y a todos los planes. Carlitos y Bartcito, you guys are the best. I love spending time with you guys, dancing, partying, cooking, doing karaoke; just every single thing was fun.

Most of my PhD years I lived in Utrecht. The crew there also played a significant role in my years in the Netherlands. Most, if not all I know about Dutch culture I owe to them. Nilito and Maraya, you two are amazing. It is very hard to find the words to describe what you two mean to me. You are my family. I have enjoyed every trip, dinner, hike, party, talk we have done. We have shared so many things, and we will have many more to come. I am very lucky to have you two in my life. Thanks for also letting me become your family. Lissetica and Jantje, thanks for always being so sweet, for caring so much about me, for always making me feel welcome. Marleentje, Sofiche, my roomies, it was great to share a house together, not only a house, but also, it has been great to spend time together. Having sgroppino nights, cooking, dancing (without Sofiche dancing), weekend trips, and festivals, everything was amazing, also with Martino. Marc and Eva, thanks for your hospitality, for making me feel at home, for caring about me. Luc is lucky to have you both as parents. Rodrigo and Laura, you guys are great, I love to see you, party with you, dance with you, even cooking; it is always a party when you both are around. Thanks for always looking after me. Joppie, Freekie, and Stevie D, the hermannus gang, it was great to have you as roomies (Stevie D counts too). Thanks for being so sweet always, for the movies, for the parties, for caring about me, but most of all, for dancing all the reggaeton with me. Ilo, Robbie and Lucie, thanks for all the amazing time together, not only the famous barbecues, but also for caring about me, always. Stein and Estherita, thanks for all the laughs, talks about helium, and parties. Keep them coming.

Ethiopia was an amazing experience, where I became “gamash habesha”, where I fell in love with the country, and met great people. At ILRI, Tadelle, thanks for your support, for opening doors to me, and for believing in my work. Ayda, Lidia, Wubie, Mickie, Agrau, Netsy, Kume, Wondy, Adebabay, Endeshaw, Tesfahun, Kelem, thanks for all your support, and always helping me with a smile. Setegn, it was great to work together. Fasilo, thanks for always being so attentive. Olivier, thanks for your positivity, interest, and support. Seblecita and Meronche, thanks for all the lunches at Zebu. Thanks Seblecita for being so sweet always, and for having the most radiant smile of all.

I also built my family outside of ILRI. Janiek, Alex, Juli, Britt, and Lolis, you made my first months in Ethiopia unforgettable. Thanks for always being around, after all these years we are still in touch, and we will be for many more years to come. You became my family; I love you all very much. It is always the greatest feeling to be together. Luisa, Lydi, Franzi, Chuchu, Nibbyje, Ghedi, Maya, Kourt, Zach, thanks for all the love, joy, parties, tibs and turbo, basically, all the moments we spent together. Flavis, te adoro, eres una mujer incredible. Caitlin and Tombrero, I love you both. Thanks for all the love and for looking after me. I am very glad you found each other (especially to build garden furniture). Clarita, thanks for being so sweet always, so generous and loving. Cass, eres como mi hermana, gracias por toda la generosidad y por todos los momentos que vivimos. Dormir en tu sofá fueron de los mejores meses de mi vida. Te adoro! Geertcito, in these years you became my brother. Thanks for always being there, for the trips, the parties, all the talks, basically, for all the moments we spent together. Thanks for not judging, for supporting me, and taking care of me.

Thanks to my family in the Netherlands, Joyce, Piet, and Luc. Thanks for always being there for me, for taking care of me, and for all the love you´ve given me all these years. Your support has been amazing. I am very lucky and feel very happy to be part of your family. Toon, mi cosis, mi amor lindo, I´m very lucky to have you in my life. Without your support, I would never have done this. Thanks for your endless love, for your respect, for giving me wings and for being there during every step. Thanks for being in my life and for making it better. I can´t wait to start living the rest of our lives together. Te amo infinito!

Colophon

The research described in this thesis was conducted through a joint project between Wageningen University & Research (WUR) and the International Livestock Research Institute (ILRI) and funded by the Koepon Foundation.

The cover of this thesis was designed by Jesse Michaël Akkerman.

Printed by: DigiForce | proefschriftmaken.nl