

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 2053*

Exploring the Genetic Landscape of Chicken Populations

*Admixture, Growth QTLs, and Long-Term Selection
Dynamics*

JEN-HSIANG OU



ACTA UNIVERSITATIS
UPSALIENSIS
2024

ISSN 1651-6206
ISBN 978-91-513-2137-0
urn:nbn:se:uu:diva-527022



UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in room A1:111a, BMC, Husargatan 3, Uppsala, Thursday, 13 June 2024 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Dominic Wright (Department of Physics, Chemistry and Biology, Linköping University).

Abstract

Ou, J.-H. 2024. Exploring the Genetic Landscape of Chicken Populations. Admixture, Growth QTLs, and Long-Term Selection Dynamics. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 2053. 36 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-2137-0.

This thesis analyzes the genetic structure of chicken populations across different breeding histories and environments. Genomic methodologies were used to uncover complex traits and domestication history over time. The work consists of three studies contributing to a broader understanding of chicken genetic diversity and the impact of selective breeding practices.

The first study delves into the global chicken population, using genome-wide analysis to uncover the intricate fine structure and historical admixture events that have shaped these populations. The research has unveiled significant connections between populations and pivotal breeding events, highlighting the complex relationships within chicken populations. This study offers intriguing insights into the genetic continuity and admixture patterns across diverse chicken breeds, from junglefowl to commercial lines.

The second study focuses on the genetic complexity within a specific quantitative trait locus (QTL) region known as *Growth1*, which is influential in chicken growth. This study, conducted using an advanced intercross line from the Virginia body weight line, identifies significant additive, haplotype, and epistasis effects within the *Growth1* QTL region. The findings challenge simplistic genetic models by demonstrating the involvement of multiple loci in regulating body weight and contribute to understanding complex trait architecture.

The third study extends the investigation to the long-term effects of selection on chicken lines, providing a deeper understanding of the genetic mechanisms underlying selection responses. By mapping multiple additive QTLs associated with body weight compared with the GWA study results, several novel regions were determined and are still contributing to the selection response even after 40 generations of intense selection.

These different views provide practical insights into chickens' intricate genetic makeup. By analyzing their domestication history, genetic variation effects, and the population's response to selective breeding, we better understand one of the most important economic organisms for humans — the chicken. This understanding can potentially inform and improve selective breeding practices, leading to more efficient and sustainable poultry production.

Keywords: Virginia Chicken Lines, population genetics, QTL, admixture, selective breeding, bioinformatics

Jen-Hsiang Ou, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

© Jen-Hsiang Ou 2024

ISSN 1651-6206

ISBN 978-91-513-2137-0

URN urn:nbn:se:uu:diva-527022 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-527022>)

For all who work with the Virginia Chicken Lines.

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I. Guo, Y., **Ou, J.-H.**, Zan, Y., Wang, Y., Li, H., Zhu, C., Chen, K., Zhou, X., Hu, X., & Carlborg, Ö. (2021). Researching on the fine structure and admixture of the worldwide chicken population reveal connections between populations and important events in breeding history. *Evolutionary Applications*, 15(4), 553–564.
- II. **Ou, J.-H.**, Rönneburg, T., Carlborg, Ö, Honaker, F. C., Siegel, P. B., & Rubin C.-J. (2024). Complex genetic architecture of the chicken *Growth1* QTL region. Accepted by Plos One.
- III. Rönneburg, T., **Ou, J.-H.**, Pettersson, M., Carlborg, Ö, Honaker, F. C., Siegel, P. B., & Rubin C.-J. (2024). Within-line segregation contributes to long-term, single-trait selection responses in the Virginia chicken lines. Manuscript.

Reprints were made with permission from the respective publishers.

Contents

| | | |
|--------|--|----|
| 1. | Introduction..... | 9 |
| 2. | Quantitative trait genetics | 11 |
| 2.1. | Polygene | 11 |
| 2.2. | The decomposition of genotypic value and genetic variance..... | 12 |
| 2.2.1. | The decomposition of genotypic value | 12 |
| 2.2.2. | The decomposition of genetic variance | 13 |
| 3. | Population genetics | 15 |
| 3.1. | Population structure analysis | 15 |
| 3.1.1. | Basic population genetic analysis | 15 |
| 3.1.2. | Principal component analysis..... | 16 |
| 3.1.3. | Model-based clustering analysis | 16 |
| 4. | The Virginia body weight lines..... | 17 |
| 5. | Comprehensive analysis of genetic variation | 19 |
| 5.1. | QTL mapping | 19 |
| 5.2. | Genome-wide association study | 19 |
| 5.3. | Haplotype-based association analysis..... | 20 |
| 5.4. | Natural and orthogonal interaction model..... | 20 |
| 5.5. | Variant annotation | 21 |
| 6. | Summary of manuscripts | 22 |
| 6.1. | Manuscript 1 | 23 |
| 6.2. | Manuscript 2 | 25 |
| 6.3. | Manuscript 3 | 27 |
| 7. | Limitations and future works | 29 |
| 7.1. | The black box of gene action..... | 29 |
| 7.2. | The quality of the data | 29 |
| 7.3. | Incomplete explanation of variance..... | 30 |
| 7.4. | Potential future works..... | 30 |
| 8. | References..... | 32 |
| | Acknowledgment | 34 |

Abbreviations

| | |
|-------|---|
| HWS | High-weight selected line |
| LWS | Low-weight selected line |
| AIL | Advanced intercross line |
| QTL | Quantitative trait locus |
| SNP | Single-Nucleotide Polymorphism |
| GWAS | Genome-wide association study |
| vGWAS | Variance-heterogeneity GWAS |
| NOIA | Natural and orthogonal interaction models |
| LD | Linkage disequilibrium |
| ROH | Run of homozygosity |
| PCA | Principal component analysis |
| PC | Principal component |
| BW8 | 8-week body weight |
| VEP | Ensembl Variant effect predictor |
| WL | White Leghorn |
| BM | Black Minorca |

1. Introduction

Individuals possess unique characteristics that we often take for granted, such as being tall or short, black or white, chubby or skinny. Such traits do not occur by chance. Instead, research has shown that many of them were inherited and passed down from generation to generation.

Researchers have been keen to uncover the factors and mechanisms of changing the phenotype for many years. Darwinian evolution posits that species evolve through natural selection, where individuals with traits better suited to their environment are more likely to survive and reproduce. Over time, these advantageous traits become more common in the population, leading to gradual changes in the species (Darwin, 1859). By simplifying the observation of the shape and color of the pea and its flower, Gregor Mendel found regular patterns in how discrete characteristics are inherited (Mendel, 1866). Later, Galton studied human traits' inheritance to address how continuous traits are inherited and describe these using statistics (Galton, 1876). R. A. Fisher considered findings from Mendel and Galton and constructed the fundamentals of statistical genetics. He observed that the phenotypic value is a sum of minor effects of many factors (now known as alleles). The decomposition of phenotypic value and variance provides chances for association studies nowadays (Fisher, 1919).

At the time, scientists did not know the molecular basis of inheritance patterns. After several decades of development, affordable and accurate sequencing platforms are available. These provide researchers with information for studying the link between molecular genetic variation and the inheritance of complex traits.

We studied chicken (*Gallus gallus*) since it is an important model organism in genetic research. Historically, chickens were among the first animals domesticated for agriculture, which provides an excellent background for studying selected breeding and domestication. Notably, chickens are economically and agriculturally crucial since they are a major protein source in their meat and eggs. Over the years, research in chickens' growth, production, feed efficiency, and disease resistance continued. The chicken genome was one of the first animal genomes to be sequenced for genetic research, and lots of information was collected for comparative genetics and evolutionary studies. The Virginia Body Weight Line is used in all the manuscripts in this thesis. Established in 1957, the Virginia Body Weight Line's base population (S_0) was

created by crossing seven inbred lines of White Plymouth Rock chickens (Dunnington & Siegel, 1996; Harrison et al., 2023; Márquez et al., 2010; Siegel, 1962). The bidirectional selection has been followed until today and has reached 67 generations. At the generation of S₄₀, the high (HWS) and low (LWS) weight selected lines displayed a nine-fold difference in average body weight at eight weeks of age. An advanced intercross line (AIL) was produced from this generation, resulting in 18 generations (F₁-F₁₈).

In the first study of my thesis, we broadly explored the global chicken population. Including genome-wide data from American, Chinese, Indonesian, and European chicken breeds. Examining junglefowl, indigenous varieties, and modern commercial lines. We found unexpected clustering patterns, historical links between breeds worldwide, and insights into the admixture events.

Following the first paper, the focus shifted to the genetic complexity of growth-related traits in one of the most studied QTL regions, the *Growth1* region. As the AIL population extended to 18 generations, we have a better resolution than previous research due to the accumulation of recombination over generations and having more samples. Including genome-wide association studies (GWAS), variance-heterogeneity GWAS (vGWAS), haplotype-based association studies, natural and orthogonal interaction models (NOIA), and publicly available annotation information, we proposed a possible interaction network in this region and highlighting non-coding regions associated with genes implicated in chicken body weight regulation.

Lastly, we studied the adaptation of the Virginia chicken line to long-term single-trait selection. The research revealed overlaps and differences between the higher-resolution QTL mapping and the GWAS. The result shows that more than a quarter of the variants are still fully segregating in both lines, indicating that they are still responding to selection even after 40 generations of intense selection. Several candidate genes located in the overlapping region of QTL and GWAS were determined and considered evolutionarily critical.

2. Quantitative trait genetics

Quantitative traits, also called complex traits, are traits in living organisms that are not simply controlled by single factors. Instead of being controlled by a single gene, as in Mendel's traits, these traits result from the accumulation of multiple minor effect genes, environmental influences, and their interactions. These cause the expression of the traits, which are often exhibited as a distribution in the population.

Studying quantitative traits has been crucial in biology, agriculture, and medicine, as it provides insight into understanding the evolutionary process, increases crop harvest, and uncovers the cause of illness. For example, in medicine, understanding the genetic basis of complex diseases such as type 2 diabetes can lead to better prevention and treatment strategies (Shi et al., 2016). Analyzing the genetic components of crop yield traits in agriculture can significantly contribute to food security by breeding more efficient and resilient crop varieties (Zhao et al., 2011).

However, analyzing quantitative traits takes work. Although many different models help us uncover the black box of how genes act, the minority effect size of genes is hard to determine. Sequencing technology development makes high-density markers affordable, while with many dozen genes, the effect ratio of being explained still needs to be improved. Despite all these challenges, researchers still have advanced in this field. Researchers have successfully identified regions, the quantitative trait loci (QTL), in the genome associated with these complex traits. Various models for association studies under different assumptions were developed to determine effects. Higher-order interactions and phenotypic influences were considered in the model to explain the phenotypic variance better. All these enlarge the peephole to view the genetic architecture better.

2.1. Polygene

As described in the previous section, polygenes play a crucial role in producing a variety of traits that can be measured in organisms. Polygenes have minor effects individually, unlike major genes, which follow simple Mendel's genetic rules that strongly influence traits. The accumulation of all effects

from polygenes produces a wide and continuous range of differences rather than an either-or type of difference.

The mix of many genes, environmental effects, and their interactions results in traits often spread across a bell curve distribution, with most individuals having measurements close to the average and fewer being more extremes. These observations move the simple view of genetics away, showing us that the combined minor effect of genes leads to diverse results.

2.2. The decomposition of genotypic value and genetic variance

The decomposition of genotypic value and genetic variance is crucial in quantitative genetics, which aims to understand the impact of genetic variation on individual phenotypes. As genetics research evolved, people realized that more comprehensive models were needed to explain traits better. Until molecular genetics developed, researchers conducted deeper research on the impact of genes and genotypes on traits, laying the foundation for quantitative genetic research. With the development accompanied by statistical methods, researchers began systematically analyzing genotype value and genetic variance.

2.2.1. The decomposition of genotypic value

The relationship between the number of an allele and genotypic value may only be linear if the alleles interact additively with all other alleles. Considering the closest linear relationship is still valuable as it enables us to estimate genotypic values based on additivity and deviations from the expected value caused by dominance. Starting from the general model of $Y = G + E$, where Y is the phenotypic value, G is the genotypic value, and E is the environment deviation. The mean environmental deviation is often considered 0, so the mean genotypic value equals the mean phenotypic value. Here, we can decompose the genotypic value into additive and dominance effects. Assuming the individuals carried allele A_i and A_j at locus A we will have:

$$G_{ij} = G_{ij}^A + \delta_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij} \quad (1)$$

Where μ_G is the genotypic mean in the population, α_i and α_j are mean additive effects for allele i and j , and δ_{ij} is the dominance effect. For a biallelic locus, the sum of two available alleles is two. Saying that having $N_1 A_1$ alleles and $N_2 A_2$ alleles, $N = N_1 + N_2 = 2$. Equation (1) can be rewritten as the following equation.

$$G_{ij} = \mu_G + \alpha_1(2 - N_2) + \alpha_2 N_2 + \delta_{ij} = \tau + (\alpha_2 - \alpha_1)N_2 + \delta_{ij} \quad (2)$$

Where $\tau = \mu_G + 2\alpha_1$ is the intercept of the linear regression and $\alpha = \alpha_1 - \alpha_2$ is the slope. We can have the genotypic predicted value with linear regression by taking the expected value for equation (2).

$$\hat{G}_{ij} = E(G_{ij}) = \mu_G + \alpha_1 E(N_1) + \alpha_2 E(N_2) + 0 \quad (3)$$

As we know, the expected value of the residual (δ) is zero, and the expected value of the number of alleles is two times its allele frequency.

$$p_1 = \frac{E(N_1)}{2}, p_2 = \frac{E(N_2)}{2} \quad (4)$$

Knowing $p_1 + p_2 = 1$ with equation (3), we have $\alpha_2 = p_1\alpha$ and $\alpha_1 = -p_2\alpha$.

2.2.2. The decomposition of genetic variance

Now, we set the reference point at the midpoint between the genotypic values of two homozygous genotypes. The genotypic value of A_1A_1 , A_1A_2 , and A_2A_2 can be assigned by $-a$, d , and a , respectively. The $k = \frac{d}{a}$ ratio measures dominance, saying that $d = k = 0$ represents complete additive and $k = -1$ or 1 implies complete dominance. The average effect of allelic substitution (α) can be denoted through this often-used reference point in the following equation.

$$\alpha = a[1 + k(p_1 - p_2)] \quad (5)$$

Recalling the regression shown by equation (1), the genetic variance can be expressed as below.

$$\begin{aligned} \text{Var}(G) &= \text{Var}(G^A + \delta) \\ &= \text{Var}(G^A) + \text{Var}(G^A, \delta) + \text{Var}(\delta) \\ &= \text{Var}(G^A) + \text{Var}(\delta) \end{aligned} \quad (6)$$

The additive genetic variance (σ_A^2) and the dominant genetic variance (σ_D^2) can be computed by allele frequency (p_1 and p_2), dominance coefficient (k), and homozygous effect (a).

$$\begin{aligned} \sigma_A^2 &= 2p_1p_2\alpha^2 \\ \sigma_D^2 &= (2p_1p_2ak)^2 \end{aligned}$$

Understanding how genotypic value and genetic variance are decomposed is essential for quantitative genetic research. By dissecting genetic effects into additive, dominance, and interaction effects, researchers gain insights into how genes work together and how they are inherited. Understanding the interplay between multiple genetic regions also empowers techniques such as QTL mapping and GWAS.

3. Population genetics

Population genetics is a field dealing with genetic principles affecting populations of organisms. These populations could be natural or experimental. As studies focus on the whole population, distribution and frequency are the main factors in discussing variants. Fundamentals of population genetics were established from Mendelian and Darwinian evolution. Factors including mutation, migration, selection, and genetic drift are critical since they contribute to the dynamic that shapes the population. Statistical models and molecular techniques are employed to analyze genetic structure, gene flow, and domesticated history to understand the population better.

This thesis applies these principles to an experimental chicken line. By examining the genomic structure, diversity, and historical events influencing populations, the research contributes to the broader view of population genetics while addressing questions related to complex traits.

3.1. Population structure analysis

Population structure analysis aims to detect genetic variation patterns and determine the existence of subpopulations within larger populations. These methods help understand the various genetic compositions of species and the genetic composition, migration patterns, historical occurrences, and adaptation.

3.1.1. Basic population genetic analysis

Genetic variation, driven by geography or evolution, is crucial for population genetic studies. In this thesis, we included several methods for better uncovering the complex genetic architecture of chickens.

Linkage disequilibrium (LD) is a widely used measure in population genetics. It shows non-random allele associations among loci. This indicates the genetic linkage between loci and how far apart they are. A higher LD score indicates that alleles are physically close or the population has experienced a bottleneck or founder event. Over generations, the accumulation of recombination causes LD to decay unless maintained by selection or genetic drift. This

provides a higher resolution for our association studies with the later generations of the chicken intercross line.

The fixation index (F_{ST}) measures population differentiation due to genetic structure. It is estimated from genetic polymorphism data. A score of 1 indicates that the population is completely different, and a score of 0 is complete panmixia.

Haplotype structures and how they change over time are also important in population genetics. Determining haplotype-centric selection is a way to detect selection events that have yet to reach fixation. This provides insights into ongoing adaptive processes.

Run of homozygosity (ROH) analysis was used to understand populations' genetic makeup. It measures the contiguous lengths of homozygous genotypes in the sample, in which identical haplotypes were inherited from parents to their offspring. The presence and extent of ROHs indicate population demographics such as bottlenecks or inbreeding.

All these methods are studied together for a better view of the complex architecture of the population.

3.1.2. Principal component analysis

Principal component analysis (PCA) is a powerful method that linearly transforms genetic data into a new coordinate system to understand population structure better. The data's greatest variance by some scalar projection lies on the first principal component (PC), the second greatest variance on the second PC, and so on. The PCs are linear combinations of effects, which allow biological researchers to figure out which factors cluster samples into groups. Meanwhile, the genotype matrix in genetic research often has an ultra-low information density. PCA provides a sufficient way of reducing dimensions without losing most information.

3.1.3. Model-based clustering analysis

Model-based clustering methods are widely used in genetic research. These methods assume that the multivariate data is generated from an admixture of component models that follow a probability distribution (Banerjee & Shan, 2011). The ADMIXTURE software estimates ancestries by driving maximum likelihood estimations in a parametric model (Alexander & Lange, 2011). Using the cross-validation method, we could determine the optimal K, which indicates the number of ancestral populations. The result provides a view of population structure and identifies individuals with mixed genetic ancestry. Relatedness patterns, subpopulations, and historical admixture events could also be discovered.

4. The Virginia body weight lines

The story starts with the domestication of chickens from the wild ancestor, the red jungle fowl, 6000 to 8000 years ago. In the 1800s, migration and admixture events created the modern chicken breeds for meat or egg production in the US. Founders of the Virginia body weight lines were established by crossing seven lines (Lillie et al., 2019). Bidirectional selection for 8-week body weight (BW8) was started in 1957 by Paul Siegel and his colleagues to study variations associated with body weight (Dunnington & Siegel, 1996; Harrison et al., 2023; Márquez et al., 2010; Siegel, 1962). The breeding process was done by segregating chickens based on their body weight, with those with higher BW8 as parents of high-weight selected (HWS) lines and those with lower BW8 as parents of low-weight selected (LWS) lines. After 40 generations of bidirectional selection, a noticeable nine-fold difference in the average BW8 between HWS and LWS lines was observed (Jacobsson et al., 2005).

The advanced intercross line (AIL) was established after extensive selection. By crossing chickens from the 41st generation of HWS and LWS lines, the intercross line spanned 17 generations (F₂-F₁₈). This intercrossing approach provides opportunities to explore the genetic architecture and how it is associated with BW8. The extension of the AIL generations accumulates recombination events and increases resolution for association studies. The breeding history of generating HWS, LWS, and AIL lines is presented in Figure 1.

It is important to note that all procedures undertaken in this population adhered to rigorous ethical standards, as outlined by the Virginia Tech Animal Care and Use Committee protocols (IACUC-15-136), ensuring the humane and responsible treatment of the research subjects.

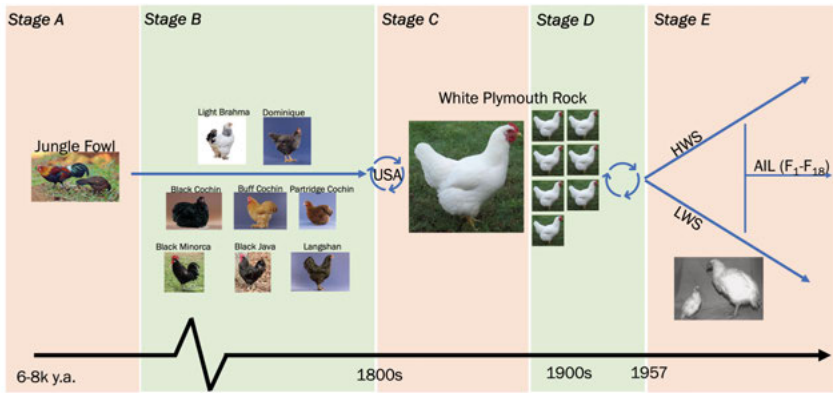


Figure 1. The establishment history of the Virginia chicken body weight lines.

5. Comprehensive analysis of genetic variation

Exploring genetic variation is crucial in genetic research. QTL mapping and GWAS are power tools for uncovering the complex relationship between genome variations and observable traits. This session will provide an overview of QTL mapping and GWAS, highlighting their distinct contributions to our understanding of genetic variation.

5.1. QTL mapping

Quantitative traits loci mapping is an approach to discovering the relationship between genotypes and phenotypes. It localizes QTL regions on the genome that affect quantitative traits in the population. Finding the correct position of causal QTL helps us better understand the genetic mechanisms (Zeng, 2001). This includes determining QTLs' additive, dominant, and epistatic effects and their possible linkage relationships with multiple traits.

Rönneburg and his colleagues first mapped QTL using the AIL population with F_2 to F_{18} generations. According to their study, the genotypes were imputed to the most likely high- and low-weight selection line founder haplotype at SNP locations across the genome. The SNP genotypes were smoothed across 1 Mb bins for genome-wide, line-cross QTL analysis in the studies. However, in addition to that, higher-resolution mapping was conducted using the founder-line blocks imputed using Stripes. This involved considering all imputed recombination breakpoints without further smoothing (Rönneburg et al., 2023). The QTL mapping result is again compared with the association study in the third paper of this thesis to identify novel regions and how they contributed to the selection response.

5.2. Genome-wide association study

GWAS is performed by identifying the association between genetic variants, often SNPs, and traits. On the principle of linkage disequilibrium, research uses a massive number of genetic markers to find those markers associated with traits that have direct or indirect linkage with causal genes. The

association is tested by comparing the variant frequencies found in individuals having different phenotypes. Unlike QTL mapping, GWAS does not rely on controlled crosses but applies to natural populations. It is useful for conducting a broad population analysis and examining genetic variations across populations.

Various GWAS models provide approaches for discovering variants. This thesis's single-marker association study is performed mainly to determine the candidate markers that could link to the causal genes. Multi-locus GWAS approaches to determining a set of markers explain the maximum amount of genotypic variances.

5.3. Haplotype-based association analysis

Haplotype-based association analysis is useful for complex trait genetic research. Instead of analyzing markers individually, haplotypes are a combination of alleles located nearby and likely to be inherited together. It increases the statistical power of the association study by several reasons. First, the resolution could be improved. As haplotypes capture more biological information than individual SNPs, it allows a more comprehensive analysis of the genetic architecture. When SNPs are in LD, the haplotype can accurately reflect the genetic variance. Second, it could capture the effect of multiple loci when multiple genetic variations contribute to a phenotype. Third, it can capture interaction effects. We can analyze haplotypes instead of SNP markers individually to detect the interactions between alleles at different loci (epistasis). And lastly, haplotype structures can significantly differ between populations due to historical events. This provides insights into the evolutionary history or migration patterns of the population.

5.4. Natural and orthogonal interaction model

The natural and orthogonal interactions (NOIA) model was developed to estimate the main and interaction effects between loci while accounting for the unbalanced allele frequency (Álvarez-Castro & Carlborg, 2007). The derivation of the model began with the following formulation and its matrix form.

$$G = S \cdot E \tag{7}$$

$$\begin{pmatrix} G_{11} \\ G_{12} \\ G_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{pmatrix} \cdot \begin{pmatrix} R \\ a \\ d \end{pmatrix} \tag{8}$$

Where the reference point $R = G_{11}$. Extended it to two loci (A and B) case having genetic effect matrix S_A and S_B , the equation (7) can be rewrite as $G_{AB} = (S_B \otimes S_A) \cdot E_{AB}$. And thus, the genetic effect vector can be calculated by $E_{AB} = (S_B^{-1} \otimes S_A^{-1}) \cdot G_{AB}$. For more general cases the derivation accounting for genotype frequencies, the genetic effect design matrix (S_F) then become:

$$S_F = \begin{pmatrix} 1 & -p_{12} - 2p_{22} & -p_{12} \\ 1 & 1 - p_{12} - 2p_{22} & 1 - p_{12} \\ 1 & 2 - p_{12} - 2p_{22} & -p_{12} \end{pmatrix} \quad (9)$$

Where p_{11} , p_{12} , and p_{22} are genotype frequencies and the reference point in this case is $R = p_{11}G_{11} + p_{12}G_{12} + p_{22}G_{22}$. For a more complex multi-locus situation, it can be obtained as an extension of this by taking the Kronecker product as describe. When $p_{11} = p_{22}$ or $p_{12} = 0$, the NOIA formula is orthogonal, which indicate that the genetic effects can be properly estimated. As the genotype frequencies vary across the loci, this model will provide better description of the interaction effects in a multi-locus model than general epistasis model (Pettersson et al., 2011).

5.5. Variant annotation

In this thesis, we use SnpEff and Ensembl Variant Effect Predictor (VEP) for annotating our genetic variants. Including variant information from different databases provide a chance for researchers better understand and priorities candidate variants. First, variant annotation helps determine the potential functional consequence such as protein structure changes. This includes whether the variant is predicted to lead to a nonsynonymous change in amino acids, alter splice sites, or gene regulatory regions. Second, for clinical research, it could help identify which mutations are pathogenic. In the chicken study, we sometimes could find genes annotated for known association with increasing risk of disease of phenotype changes. Third, we could annotate variants by how conserved they have been kept throughout evolution. Such as the conservation scores used in our study which provide a view of whether the marker is evolutionary important. Finally, with all the additional information provided, we could sort all variants by the likelihood of them being an important candidate. In genetic association studies are often faced with a huge number of candidate variants and being able to prioritize variants saves time and resources.

6. Summary of manuscripts

The manuscripts in this thesis delve into the intricate genetic landscapes of chicken populations, focusing on their origins, genomic structures, and the complex genetic architecture governing growth-related traits. Through a comprehensive examination of diverse chicken (*Gallus gallus*) breeds, these studies employ genomic techniques to unravel the relationships, admixture events, and selection pressures that have shaped the genetic diversity of these avian populations.

Manuscript 1 broadly explores the global chicken population, utilizing genome-wide analyses and multiple methodologies to reveal connections between populations and historical breeding events. The study spans American, Chinese, Indonesian, and European chicken breeds, examining junglefowl, indigenous varieties, and modern commercial lines. Noteworthy findings include unexpected clustering patterns, historical links between Chinese and Rhode Island Red breeds, and insights into admixture events, shedding light on the intricate relationships within chicken populations.

The Virginia body weight line, included in Manuscript 1, serves as the primary dataset for Manuscript 2 and 3. Established in 1957, the Virginia body weight line's base population (S_0) was created through the crossing of seven predominantly inbred lines of White Plymouth Rock chicken. The bidirectional selection was done from the S_0 generation for a single trait, 56-day body weight. Following 40 generations of selective breeding, the high (HWS) and low (LWS) weight lines exhibited a 9-fold difference in average body weight. Some HWS and LWS chickens from the S_{41} generation of the bidirectional selection line produced an advanced intercross line (AIL), resulting in 18 generations (F_1 - F_{18}).

In Manuscript 2, the focus shifts to the genetic complexity of growth-related traits, explicitly exploring the extended Growth1 quantitative trait locus (QTL) region in Virginia body weight lines. Despite previous efforts, understanding growth-related QTLs remains limited. An AIL with 18 generations is introduced for enhanced resolution. The genome-wide association study (GWAS) of the extended Growth1 QTL region identifies significant additive effect signals, and a variance-heterogeneity GWAS (vGWAS) reveals variance effect signals in this region. Haplotype and interaction analyses uncover intricate genetic makeup and emphasize the involvement of multiple loci in regulating body weight. The study explores independent marker effects,

incorporating a Natural and Orthogonal Interaction (NOIA) model, and identifies significant markers and epistasis effects. Annotation of variants highlights non-coding regions associated with genes implicated in chicken body weight regulation. This study challenges simplistic views of polygenic traits and provides a detailed understanding of chickens' genetic mechanisms governing growth.

Manuscript 3 further contributes to understanding genetic adaptations in the Virginia chicken lines by explicitly examining the long-term effects of single-trait selection responses. By mapping 42 additive QTLs associated with 56-day body weight, the research reveals overlaps and exposes differences between higher-resolution QTL analysis and the GWAS. The findings reveal independent signals beyond mapped QTLs, emphasizing the importance of segregating variants within and across lines and providing new insights into the dynamics of long-term selection adaptation.

These manuscripts underscore the importance of genomic approaches in unraveling the complex genetic underpinnings of chicken populations. These studies contribute to a comprehensive understanding of the evolutionary history and adaptive processes that have shaped chicken breeds by investigating relationships, admixture events, and the genetic architecture of growth-related traits.

6.1. Manuscript 1

Researching on the fine structure and admixture of the worldwide chicken population reveal connections between populations and important events in breeding history.

In this study, we analyzed 636 genomes from 43 American, Chinese, Indonesian, and European populations. Evaluated populations include junglefowl, rural indigenous chickens, and modern commercial lines bred for efficient meat and egg production.

Principal component analysis (PCA) on genome-wide SNP data shows a close relationship among most populations. Notable outliers included experimental White Plymouth Rock lines (HWS and LWS from the Virginia body weight lines), White Leghorn (WL), and Black Minorca (BM). Unexpectedly, meat-type birds (HWS, LWS) clustered with WL and BM, possibly due to ancestral contributions. Populations from Tibet and Sichuan exhibited scattered relationships, suggesting common admixture. A historical link was observed between the Chinese breed Liyang and Rhode Island Red, aligning with records of Chinese breeds influencing the creation of Rhode Island Red.

Model-based clustering is utilized to complement the PCA analysis to investigate admixture events. A K value of 8 was chosen based on minimal cross-validation error. Admixture analyses revealed significant Green

Junglefowl introgression in Indonesia's sampled Red Junglefowl population. Indigenous Chinese breeds exhibited similar population structure, with Tibetan breeds showing high admixture and close relation to Red Junglefowl. Agriculturally developed breeds clustered separately, reflecting intensive breeding and common ancestors with Asian breeds, as expected from historical records. Liyang, a descendant of the Shanghai chicken, played a significant role in cross-breeding. In the heavy body weight group, Dominique was related to various heavy Asian and Western chicken breeds. In contrast, Cobb, a modern western broiler breed, shared a majority of ancestry sources with other heavy breeds. Langshan populations and different Cochin varieties also showed close relationships.

Linkage disequilibrium (LD) and autozygosity in populations were investigated to provide insights into their demographic histories. LD patterns, assessed using pairwise SNV distributions, aligned with expectations based on known population characteristics. Populations with slower LD decay were more recently founded and often subjected to intense artificial selection in a smaller number of individuals, exemplified by the Virginia body weight lines. Autozygosity analysis, measured through runs of homozygosity (ROH), revealed the Green Junglefowl and Virginia BW lines had the largest sum of ROH, indicating historical bottlenecks and intensive selection. Modern US breeds generally exhibited higher autozygosity, consistent with smaller founding populations. Chinese local breeds showed lower genomic autozygosity, possibly due to weaker selection and larger effective population sizes. Commercial lines displayed varied ROH distributions, suggesting diverse founder events and breeding strategies.

Evolutionary relationships were studied among chicken populations through genome-wide analyses, phylogenetic approaches, and gene-centered investigations. Finestructure analysis confirmed what was known and suggested new relationships between modern and local breeds, emphasizing shared ancestry among closely clustered populations. Phylogenetic trees corroborated expected relationships and revealed novel connections, such as those between Liyang and Rhode Island Red and White Leghorn and Black Minorca. Analysis of genetic admixture using Globetrotter identified Black Minorca as a major ancestor of White Leghorn, supporting historical records. Selection signatures in pigmentation genes, like *TYR* and *MC1R*, were identified, reflecting the impact of selective breeding on coat color. Surprisingly, *ERBB3*, not *PMEL17*, was associated with the dominant white trait, challenging previous findings. The study provides a comprehensive understanding of the genetic relationships, admixture events, and selection pressures shaping the diversity of chicken breeds.

This study concludes that genomic approaches provide valuable insights into breed origins, contributions, and demographic histories, shedding light on the intricate relationships and adaptations in chicken populations.

6.2. Manuscript 2

*Complex genetic architecture of the chicken *Growth1* QTL region.*

The genetic complexity of polygenic traits represents a captivating and intricate facet of biological inheritance. As a widely used economically and scientifically important model organism, chicken provides a great model for deep into discovering the complexity of genetic architecture. Dr. Paul Siegel's team at Virginia Polytechnic Institute established the Virginia body weight lines, which bidirectionally selected chickens by body weight for more than 40 generations. Despite identifying growth-related QTLs, previous studies explain only a fraction of the variance, indicating the complexity of genetic influences. An advanced intercross line (AIL) was established to enhance resolution, producing 18 generations of intercross population (F₁-F₁₈).

In the GWAS of the *Growth1* QTL region, two significant peaks (gga1_168m and gga1_171m) were identified. When considering gga1_171m as a covariate, the significance of the left-hand peak decreased but did not disappear; however, including both peaks as covariates eliminated significant peaks in the region. Two potential explanations for this complex genetic architecture were proposed. Firstly, distinct haplotype effects might exist due to linkage disequilibrium between functional alleles not adequately captured by individual SNP markers. The LD analysis revealed that gga1_171m and gga1_168m markers are not strongly linked. Secondly, interactions between loci could contribute to nonadditive genetic variance, which may be reflected as genetic variance heterogeneity in the model.

To explore the intricate genetic makeup of the *Growth1* QTL, a variance-heterogeneity GWAS was conducted to identify SNP markers carrying variance effects. This analysis revealed four additional signals downstream of standard GWAS peaks. The selected SNP markers include gga1_171v, gga1_172v, gga1_174v, and gga1_178v.

The haplotype effect was explored as a potential explanation for the complex genetic architecture of the *Growth1* QTL. LD analysis revealed that individual SNP markers did not capture the LD between functional alleles, indicating the involvement of multiple loci in the body weight trait. Haplotype-based GWAS confirmed that the two detected GWAS peaks had independent effects, with a region of non-association between them, suggesting they are less likely to exist on the same haplotype. Ancestral haplotype association studies supported these findings, showing strong significance for the two main GWAS peaks. Analysis of HWS and LWS samples painted by ancestry information revealed fixed haplotypes in the LWS samples for the right peak, consistent with earlier studies on haplotype complexity. For the left peak (gga1_168m), three major haplotypes were identified, with the highest frequency haplotype present in both HWS and LWS samples but with a higher

frequency in LWS. The other two haplotypes existed only in either HWS or LWS samples, with significantly different effects on body weight.

Independent marker effects were assessed through stepwise selection across determined SNP markers, with sex and generation considered fixed effects. The final model retained *gga1_168m*, *gga1_171v*, and *gga1_178v* as significant markers. Evaluation of mean effects with the number of reference or alternative alleles showed that, for *gga1_168m*, *gga1_171m*, *gga1_174v*, and *gga1_178v*, body weight decreased with increasing numbers of alternative alleles. The effect size for *gga1_168m* was relatively large, though the group sizes were unbalanced. In contrast, for *gga1_171m* and *gga1_174v*, the numbers of individuals in opposite homozygote groups were more balanced. However, for *gga1_171v* and *gga1_172v*, average body weight increased with the number of alternative alleles. The least significant difference analysis was employed to test for significant differences in body weight among genotype groups for each locus.

The natural and orthogonal interaction model, incorporating six markers from GWAS and vGWAS results, considered all additive, dominant, and second-order interaction effects. Sex and generation effects were normalized within each sex-generation group. Three markers, *gga1_168m*, *gga1_174v*, and *gga1_178v*, displayed highly significant interaction effects in the upper triangle of the network, emphasizing their importance. Additionally, *gga1_171m* and *gga1_172v* showed significant interactions with *gga1_174v* and *gga1_168m*, respectively. The last marker, *gga1_171v*, exhibited a mild additive interaction effect with *gga1_178v*. The significant interactions highlighted in the model emphasize that regulating body weight by the *Growth1* QTL involves more than independent effects of individual loci.

The average and standard deviation of body weight, grouped by genotypes of two selected markers, illustrate how mean and variance change with the interaction between these markers. Association analysis, observed under different genotype conditions, revealed notable changes. In the case of the two GWAS peaks, their strength persisted in individuals carrying at least one *gga1_178v* reference allele, which is more frequent in HWS samples. However, both signals were absent in alternative allele homozygotes at *gga1_178v*. Similar observations were made when grouping samples by the genotype of the top SNP marker and *gga1_178v*, where body weight exhibited a minor difference between alternative allele homozygotes at *gga1_178v* but a significant difference for those carrying at least one reference allele. Conversely, markers *gga1_172v* and *gga1_174v* showed a different pattern, eliminating two GWAS peaks in both homozygous groups.

Variants were annotated using snpEff (GRCg6a.105), and the analysis included intersection with vertebrate PhyloP scores to prioritize variants in non-coding regions based on evolutionary constraints. Notably, SNP markers most strongly associated with body weight did not predict changes in coding regions of genes. Instead, they clustered in introns of or intergenic regions to

genes such as *Ecto-NADPH Oxidase Disulfide-Thiol Exchanger 1 (ENOX1)*, *ENSGALG00000050514*, *ENSGALG00000052226*, and *ENSGALG00000053256*. Of particular interest, *ENSGALG00000053256*, identified as a novel long non-coding RNA, has been previously implicated as a candidate gene for regulating chicken body weight. Furthermore, some strongly associated SNP variants were predicted to cause gene amino acid substitutions. Notably, the protein encoded by *TNF superfamily member 11 (TNFSF11)* has known effects on bone growth, and *ribonuclease H2 subunit B (RNASEH2B)* has previously emerged as a candidate gene for regulating beak size and shapes in Darwin's finches and overlaps a QTL for growth in an intercross between fast-growing broiler and slow-growing Chinese indigenous breeds.

6.3. Manuscript 3

Within-line segregation as contributors to long-term, single-trait selection responses in the Virginia chicken lines.

Genetic variation enables population evolution in response to selection, which has been crucial in agriculture for millennia. Studying the inheritance of complex traits, primarily through experimental crosses like the F₂ generation, has been a longstanding pursuit. While F₂ crosses historically offered advantages, modern genotyping technologies reveal limitations in resolution. Transforming F₂ into intercross lines, accumulating recombination, and exploring diverse populations address these challenges. Artificial selection, exemplified by intense selection pressure, is a testbed for understanding selection responses. Utilizing intercross lines derived from selection experiments aids in dissecting quantitative traits and understanding their dynamics. This study uses the same chicken experimental population as Manuscript 2. Contrary to assumptions of fixed variants, a more extensive dataset reveals the importance of segregating variants within and across lines, providing new insights into long-term selection adaptation and maintenance of genetic variance.

Forty-two additive QTLs associated with 56-day body weight were mapped. The results show an overlap between higher-resolution QTL analysis and the GWAS for single-locus additive effects. The GWAS-multi-locus analysis detected 41 markers after backward elimination, and 13 of the 42 QTLs were identified through the line-cross analysis. The study also revealed 24 GWAS regions that did not overlap with QTLs. The findings highlight body weight's complex genetic architecture and independent signals beyond the mapped QTLs.

The analysis revealed that most GWAS regions are represented by a single marker, except for two exceptions that are highly linked. Additionally, two QTLs were found to deconvolute into multiple independent GWAS signals.

Regarding the extent of GWAS regions, determined by linkage disequilibrium, it was observed that these regions can range up to 3.624 Mb. Interestingly, areas overlapping previously identified QTLs tend to have a slightly higher mean extent than those without a corresponding QTL peak.

There is no significant difference in F_{ST} and Minor Allele Frequency between regions overlapping previous QTLs and those that do not. Notably, a substantial fraction of novel GWAS regions still contributes to the selection response, with markers fixed or segregating in the divergent lines.

When partitioning the explained variance, markers retained after forward selection were found to explain 23.9% of the phenotypic variance. About 27.8% of the variance is attributed to markers still fully segregating in both lines, while 72.2% is contributed by markers fixed in at least one line. Moreover, GWAS regions overlapping QTLs explain 37.4% of the total variance, and GWAS regions not previously identified contribute 62.6%.

Exploring novel GWAS-only regions, it was observed that 37.8% of the weight explained is due to markers still segregating within both lines. Notably, 62.1% of the described weight is attributed to markers not previously identified but fixed in at least one line, suggesting a significant contribution to the selection response.

Missense variants were considered in analyzing candidate causal genes within the overlapping region of GWAS and QTL. Notably, the *ribonuclease H2 subunit B (RNASEH2B)* gene emerged as a potential candidate due to its significance in a previous chicken intercross line association study and its association with beak size and shape in Darwin's finches. Additionally, *GALNT7*, *TOM1*, and other genes were assessed, with SIFT predictions suggesting them to be tolerated. In contrast, *ENSGALG00000012791* and *CHD7* were identified as potentially deleterious. *ENSGALG00000012791* is linked to *thromboxane A synthase 1 (TBXAS1)* associated with certain diseases, while *CHD7* exhibits strong conservation across avian and reptilian species, indicating its evolutionary importance. Despite the high delta allele frequency and potential functional significance of an SNP in *CHD7*, no functional experiments have validated this hypothesis to date. Furthermore, the study considered high-impact variants likely to cause disruptive effects on proteins, annotated as HIGH by Ensembl variant effect predictor (VEP). Among these, the marker in the *contactin-associated protein family member 5 (CNTNAP5)* gene showed the strongest correlation with body weight. However, conservation scores did not provide evolutionary indications of functional indispensability for any markers.

In summary, this research unveils the complexity of the genetic architecture of body weight in chickens, even after 40 generations of intense selection. The findings contribute to understanding how selection has shaped genetic architecture in different lines and provide valuable insights for investigating quantitative traits across populations.

7. Limitations and future works

As we dig deeper into the world of complex genetic traits with great tools, we are still facing challenges to achieve. On one hand, the complex interaction between genes and environments is still under study. This makes it difficult to clearly identify genetic effects, especially for quantitative trait research with minor effect sizes. As in the data set (Virginia chicken lines) we used in the study, we tried to remove the batch and sex effects. These effects are often much greater than the variant effects since most QTLs have minor effects. In this case, when the environmental effects interact with the genes, the dominance of these effects could make it hard to determine the pure gene or gene-by-gene effects. On the other hand, the genetic model used today is still limited. So far, we have yet to fully explain all observed variance and the biological functions of markers. Some hope comes with advanced computing power, complex genetic models, diverse data sets, and novel methodologies.

7.1. The black box of gene action

The “black box” for genetic research is the challenge in studies referring to the incomplete understanding of the mechanisms where genes influence the expression of complex traits. Despite the advancements in genetics and genomics over the decades, understanding the mechanisms by which genes interact with each other and the environment to shape phenotypic variation is still challenging. This is especially pronounced in the study of quantitative traits influenced by polygenes, which each contribute a small effect, and the environment can affect their expression.

7.2. The quality of the data

During the publication process, journal reviewers challenged the genotype imputation quality of our AIL population. Luckily, we can achieve a good imputation result with the information of pedigree and deep-sequenced founders. This was confirmed by a high agreement between the GoldenGate assay genotype and the imputed genotype in the AIL population's early and later generations. This again reminds us of the importance of data quality. We often need

to balance cost and quality, which is not easy. Performing a well-planned experimental design and consulting with field experts and statisticians beforehand can reduce frustration. Biological data is subject to biological and technical variability, which sometimes makes it difficult for researchers to draw proper conclusions if adequate care has yet to be taken in study design and implementation. Increasing the sample size appropriately and fairly is always the right choice where possible.

7.3. Incomplete explanation of variance

Current genetic models and analyses cannot fully capture all genetic variance observed in genetic studies of complex traits like chicken body weight. Despite using methods such as QTL mapping or GWAS, much phenotypic variance still needs to be explored. This limitation could be caused by genetic interactions, rare genetic variants, and environmental factors. To uncover the complexity of genetic architecture, we should include more analytical techniques and multi-omics data. By digging different sides of the mystery, we could have a more comprehensive understanding of complex traits. We need a better resolution of genomic data and genetic analysis to identify numerous minor effect genes. This could be achieved by driving a more balanced and larger dataset with better models. Additionally, incorporating gene-environment interactions and rare genetic variants through improved models will provide a more complete explanation of variance.

7.4. Potential future works

The genetic underpinnings of complex traits are still being explored. As discussed in previous sections, many limitations still need to be addressed. Some could be done with our current tools, and some require creating new weapons for future battles.

1. Functional genomics of growth traits: Investigate the functional genomics underlying growth traits, employing RNA sequencing and other functional genomics assays to determine causal variants better and understand regulatory mechanisms.
2. Epigenetic regulation in chicken development: Study the epigenetic changes during chicken development, focusing on how these changes influence gene expression and trait manifestation.
3. Gene-environment interactions in phenotypic traits: GWAS and environmental data analyzed together to explore how environmental factors interact with the genetic factors of chickens and how they influence phenotypic traits.

4. The role of epigenetics in trait variability: Investigate how epigenetics contributes to phenotypic variation and adaptation in populations, distinguishing between genetic and epigenetic influences on trait variation.
5. Ancient DNA and population histories: Ancient DNA is used to reconstruct population histories and migrations, focusing on how historical movements have shaped current genetic diversity and structure.
6. Machine learning applications in population genetics: Develop machine learning models to predict population genetic parameters and to analyze complex datasets efficiently.
7. Integrative models of gene flow and selection: Create integrated models that combine gene flow, genetic drift, and natural selection to predict allele frequency changes and understand the interplay between migration and adaptation.
8. Statistical tools for heterogeneous data: Develop new statistical tools to analyze heterogeneous genetic data, such as combining traditional population genetic data with functional genomic data, to better understand the genetic basis of adaptation.

8. References

- Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. <https://doi.org/10.1186/1471-2105-12-246>
- Álvarez-Castro, J. M., & Carlborg, Ö. (2007). A Unified Model for Functional and Statistical Epistasis and Its Application in Quantitative Trait Loci Analysis. *Genetics*, *176*(2), 1151–1167. <https://doi.org/10.1534/genetics.106.067348>
- Banerjee, A., & Shan, H. (2011). Model-Based Clustering. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 686–689). Springer US. https://doi.org/10.1007/978-0-387-30164-8_554
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*.
- Dunnington, E. A., & Siegel, P. B. (1996). Long-Term Divergent Selection for Eight-Week Body Weight in White Plymouth Rock Chickens. *Poultry Science*, *75*(10), 1168–1179. <https://doi.org/10.3382/ps.0751168>
- Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. <https://doi.org/10.1017/S0080456800012163>
- Galton, F. (1876). A Theory of Heredity. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *5*, 329. <https://doi.org/10.2307/2840896>
- Harrison, S. J., Siegel, P. B., Honaker, C. F., & Lewis, R. M. (2023). Population dynamics of a long-term selection experiment in White Plymouth Rock chickens selected for low or high body weight. *Poultry Science*, *102*(5), 102575. <https://doi.org/10.1016/j.psj.2023.102575>
- Jacobsson, L., Park, H.-B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, P. B., & Andersson, L. (2005). Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. *Genetical Research*, *86*(2), 115–125. <https://doi.org/10.1017/S0016672305007767>
- Lillie, M., Honaker, C. F., Siegel, P. B., & Carlborg, Ö. (2019). Bidirectional Selection for Body Weight on Standing Genetic Variation in a Chicken Model. *G3 Genes|Genomes|Genetics*, *9*(4), 1165–1173. <https://doi.org/10.1534/g3.119.400038>
- Márquez, G. C., Siegel, P. B., & Lewis, R. M. (2010). Genetic diversity and population structure in lines of chickens divergently selected for high and low 8-week body weight. *Poultry Science*, *89*(12), 2580–2588. <https://doi.org/10.3382/ps.2010-01034>
- Mendel, G. (1866). *Versuche über Pflanzenhybriden*. Im Verlage des Vereines.
- Pettersson, M., Besnier, F., Siegel, P. B., & Carlborg, Ö. (2011). Replication and Explorations of High-Order Epistasis Using a Large Advanced Intercross Line Pedigree. *PLoS Genetics*, *7*(7), e1002180. <https://doi.org/10.1371/journal.pgen.1002180>

- Rönneburg, T., Zan, Y., Honaker, C. F., Siegel, P. B., & Carlborg, Ö. (2023). Low-coverage sequencing in a deep intercross of the Virginia body weight lines provides insight to the polygenic genetic architecture of growth: Novel loci revealed by increased power and improved genome-coverage. *Poultry Science*, *102*(5), 102203. <https://doi.org/10.1016/j.psj.2022.102203>
- Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *The American Journal of Human Genetics*, *99*(1), 139–153. <https://doi.org/10.1016/j.ajhg.2016.05.013>
- Siegel, P. B. (1962). Selection for Body Weight at Eight Weeks of Age. *Poultry Science*, *41*(3), 954–962. <https://doi.org/10.3382/ps.0410954>
- Zeng, Z.-B. (2001). QTL Mapping. In *Brenner's Encyclopedia of Genetics* (pp. 8–12). Elsevier. <https://doi.org/10.1016/B978-0-12-374984-0.01248-1>
- Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., Norton, G. J., Islam, M. R., Reynolds, A., Mezey, J., McClung, A. M., Bustamante, C. D., & McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications*, *2*(1), 467. <https://doi.org/10.1038/ncomms1467>

Acknowledgment

I am immensely grateful for the support and encouragement I received throughout my doctoral studies and the writing of this thesis. This journey was not easy; it was fraught with challenges and moments when the path ahead seemed overwhelmingly unclear. In times of despair, it felt like angelic hands were consistently extended to lift me. I deeply value the guidance and experiences shared by teachers, the encouragement, and even the complaints with my peers, as well as the inspiring support from friends and family outside the academic circle. I hold all this help close to my heart, so much so that words can scarcely express my full emotions.

Throughout these years of study, my seniors constantly reminded me that the true essence of pursuing a Ph.D. goes far beyond academic achievements alone. It is the journey itself where real learning occurs. As I am about to complete my thesis writing, I realize how magnificent the scenery along the way has been. I have learned to face setbacks, cope with loss, collaborate as a team, listen to others, and express my opinions. No matter where I go from here, I believe these lessons will never be in vain. As I pause here, I want to take a special moment to thank specific individuals. Without you, none of this would have been possible.

Carl, when even I began to entertain thoughts of giving up, you stepped in unhesitatingly and became the beacon of light in the latter half of my doctoral journey. I could have improved at meticulous planning, but you guided me, step by step, in crafting a viable plan and ensured that I followed through. I do not doubt that without your help, I could not have accomplished all this. From you, I learned how to approach research methodically and deeply explore the core of truth from multiple angles. These lessons will be invaluable to me in the future.

Örjan, it was you who instilled in me both confidence and opportunity. When I first joined the lab with an incomplete understanding of molecular biology, you offered me unparalleled encouragement and assistance. You designed a doctoral research topic suited for me and provided ample opportunities to experiment. From you, I learned how to make complex knowledge understandable to others and make dry scientific topics engaging, and you taught me the importance of sometimes taking a step back to see the broader picture.

Tilman, from the moment you welcomed me into your shared house and saved me from homelessness, you became a brother I will forever respect. You consistently approach everyone with the brightest attitude and unmatched generosity and are always willing to lend a hand. Over the years, we have faced all challenges in the lab together: teaching side by side, guiding each other in our research, jointly tackling computer viruses attacking our servers, rescuing precious sequencing data, trying to moderate our coffee consumption, and encouraging each other not to slack off too much (though with debatable success). I am confident that you will become an outstanding scientist, as your love for helping others will inevitably prove to be your greatest strength.

Ying, you were the guiding light who took me under your wing at the start of my doctoral journey, introducing me to the field and teaching me step-by-step. Your help extended beyond academia, assisting me with practical matters such as setting up a bank account and applying for a BankID. Even navigating the administrative processes would have been a challenge without your support.

Thibaut and **Yunzhou**, we have spent many lunchtimes and Fika breaks together, sharing the trivialities of daily life and offering each other valuable feedback during lab meetings. I am genuinely grateful for these beautiful moments we've shared.

Tanel, **Veronica**, **Rehné**, **Malin**, and **Malin**, thank you for your invaluable advice and assistance with administrative and academic matters throughout my time at school. Your reminder emails have rescued me countless times, like a horse reined in at the brink of a cliff, preventing many a forgetful mishap. I am deeply appreciative of your support.

Paul. I am profoundly grateful for the opportunity to utilize Virginia chicken body weight lines, a monumental resource shaped by the dedication of Paul and all the collaborators who have contributed to its development. Your collective efforts and perseverance have made this remarkable achievement a reality. I am incredibly fortunate to stand on the tower you have built, humbly exploring the beauty of genetics, a privilege that did not come quickly. If possible, I aspire to emulate the respected Paul, finding my passion in life and devoting myself entirely to the pursuit of elusive truths, just as pioneers like you have illuminated the path to today's brilliance.

Christa, I am grateful for your valuable and substantial suggestions while revising our research articles. Your insights have been beneficial and greatly enhanced the precision of the research paper.

To everyone in the **D11:3** corridor, the three years of the pandemic made it feel like we had just met before it was time to part ways, making our brief encounters all the more precious. I sincerely thank each of you for the conversations, shared lunches, Fika breaks, and various activities we enjoyed. I have learned about cultures worldwide and have been immensely inspired by your experiences. The insights you shared are treasured and will undoubtedly guide my decisions.

Liao and **Tung** from NTU, I am truly thankful for your invaluable guidance and mentorship during my master's studies and the benefits I continue to reap. Whenever I return to the university and converse with you, your experience and encouragement give me warmth and strength, propelling me forward courageously. Having you as my mentor as I navigate the academic world is an immense honor.

I am profoundly grateful to PoYa and her cheerful minions, aka the **class B01 European Ph.D. Students** Association from the Department of Agronomy at NTU. Growing together in such a competitive class was both challenging and joyful. Having so many classmates striving to complete their doctoral studies across various European countries has been unforgettable. We shared our struggles and vented our frustrations, and the successive graduations of our peers perfectly motivated me to quicken my pace.

Finally, I thank **all my beloved in Taiwan** who have given me the most substantial support. Your unwavering faith has allowed me to pursue my goals fearlessly and without reservation. I often worried you throughout this journey, and I am genuinely sorry. I am deeply grateful for your prayers and trust in my ability to overcome challenges. Today, I can proudly tell you that your trust was not misplaced. While my achievements may not be perfect, they mark a significant chapter in my life.

Completing my doctoral studies involved countless forms of assistance, and I can't list everyone individually here. I am fully aware that if I were to do so, the length of these acknowledgments would far surpass that of the thesis itself. However, this does not mean that your contributions are any less critical. On the contrary, each person who has offered me a helping hand or a smile holds a permanent place in my heart, and this gratitude will never fade. Through this experience, I have profoundly realized how difficult completing this thesis has been—not merely due to the academic challenges but also because it required the collective wisdom and effort of many. I am humbly grateful for all your help.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 2053

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-527022



ACTA UNIVERSITATIS
UPSALIENSIS
2024