

# Genetic Analysis of Autoimmune and Metabolic Traits in Chickens

Weronica Ek

*Faculty of Veterinary Medicine and Animal Science  
Department of Animal Breeding and Genetics  
Uppsala*

Doctoral thesis  
Swedish University of Agricultural Sciences  
Uppsala 2011

Acta Universitatis agriculturae Sueciae  
2011:73

Cover: Ronald Nelson, SLU

ISSN 1652-6880  
ISBN 978-91-576-7617-7  
© 2011 Weronica Ek, Uppsala  
Print: SLU Service/Repro, Uppsala 2011

# Genetic Analysis of Autoimmune and Metabolic Traits in Chickens

## Abstract

Studying how changes in DNA affect organisms and gene functions, can give us an increased understanding of the genetic regulation of diseases and metabolic traits. Chickens and other farm animals have long been used as a tool to study human diseases and other complex traits since they provide an excellent opportunity to study these traits in a controlled environment. In this thesis I have applied new statistical methods to experimental data from two outbred chicken crosses to investigate the genetic architecture and regulation of autoimmune and metabolic traits. The first cross was generated from two chicken lines divergently selected for high and low body weight and was used to study how a four-QTL network affects metabolic traits (papers I-III). The second cross was generated to study the genetic regulation of autoimmune systemic sclerosis (paper IV).

Our explorations of the effects of the four-QTL network on other metabolic traits indicate that the major contribution of the network is to increase body size through an overall increase in growth of all organs rather than by changing the relative size of different components. To be able to study the networks affect on metabolic traits further, we first generated and analysed an advanced intercross line to narrow down the regions. Once the regions were fine mapped, we designed and generated a so called introgression line. Three of the fine mapped QTL were transferred (introgressed) from the low weight line into a high weight line background. This introgression line can be used to further investigate the QTL network. In the second cross, three suggestive autosomal susceptibility QTL were mapped in the chicken genome, with several orthologues of human systemic sclerosis candidate genes located within two of these loci. We also observed less disease in females with a W chromosome inherited from the diseased founder line.

*Keywords:* chicken, metabolic traits, body weight, QTL, introgression, epistasis, scleroderma, UCD-200, AIL

*Author's address:* Weronica Ek, SLU, Department of Animal Breeding and Genetics, Box, 7023, 750 07 Uppsala, Sweden

*E-mail:* Weronica.Ek@slu.se

*Dedicated to my mother and Jenny...*

*Ju mer man tänker, desto mer inser man att det inte finns  
något enkelt svar...  
/ A.A Milnes*

# Contents

List of Publications	7
Abbreviations	8
Introduction	9
Background	13
Metabolic traits	13
Autoimmune diseases	13
Systemic sclerosis	14
The chicken as an animal model	14
Chicken model for metabolic traits	15
Chicken model for human autoimmune systemic sclerosis	16
The chicken genome	17
Genetic linkage maps	18
Genetic markers	18
Genetic linkage	19
Using recombination to construct a genetic linkage map	21
Introduction to QTL mapping	21
Experimental crosses in QTL mapping	23
QTL mapping using outbred crosses	24
F <sub>2</sub> and backcross design	25
Fine mapping using an advanced intercross line	27
Introgression lines	27
QTL mapping methods used in this thesis	28
QTL mapping by linear regression	28
Variance component QTL mapping	30
Epistatic QTL mapping	31
Aims of the thesis	33
Summary of investigations	35
Analysing metabolic traits in chickens (papers I-III)	35
Paper I. Genetic analysis of metabolic traits in an intercross between body weight selected chickens	35
Methods	36
Results and discussion	37

Paper II. Fine mapping and replication of QTL in an outbred advanced intercross line	39
Methods	39
Results and discussion	40
Paper III. Using marker assisted introgression to replicate multiple epistatic loci with large effect on juvenile body weight in chicken	41
Methods	41
Results and discussion	42
Paper IV. Mapping systemic sclerosis genes in an outbred chicken intercross	45
Methods	45
<i>Linkage map</i>	46
<i>Generalised linear regression</i>	46
<i>Variance component approach</i>	47
<i>Epistatic analysis</i>	47
Results and discussion	47
Future research	49
Ongoing work	49
Metabolic traits	49
Autoimmune traits	49
Near future	50
Looking ahead	51
Conclusions	53
Paper I	53
Paper II	53
Paper III	53
Paper IV	54
Genetisk analys av autoimmuna och metabola egenskaper hos kyckling	55
Bakgrund	56
Sammanfattning av studierna	57
Metabola egenskaper (artikel I-III)	57
Autoimmuna egenskaper (artikel IV)	58
References	61
Acknowledgements	67

## List of Publications

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

- I **Weronica Ek**, Lina Strömstedt, Per Wahlberg, Paul Siegel, Leif Andersson & Örjan Carlborg (2010). Genetic analysis of metabolic traits in an intercross between body weight selected chicken lines. *Physiol Genomics* 42;20-22. doi:10.1152/physiolgenomics.00149.2009
- II Francois Besnier, Per Wahlberg, Lars Rönnegård, **Weronica Ek**, Leif Andersson, Paul B Siegel & Örjan Carlborg (2010). Fine mapping and replication of QTL in outbred Advance Intercross Line. *Genetics Selection Evolution* 2011;43:3. BioMed Central
- III **Weronica Ek**, Stefan Marklund, Ashok Ragawand, Paul Siegel, William Muir & Örjan Carlborg (2011). Generation of a multi-locus chicken introgression line to study the effects of genetic interactions on metabolic phenotypes in chicken. *Manuscript*.
- IV **Weronica Ek**, Anna-Stina Sahlqvist, Lucy Crooks, Roswitha Sgonc, Hermann Dietrich, George Wick, Olov Ekvall, Leif Andersson, Örjan Carlborg, Olle Kämpe & Susanne Kerje (2011). Mapping Scleroderma genes in an outbred chicken intercross. *Submitted*.

Reprints were made with permission from the publishers.

## Abbreviations

QTL	Quantitative trait locus
UCD-200	University of California at Davis chicken line 200
RJF	Red jungle fowl
LM	Linear model
GLM	Generalised linear model
VC	Variance component model
GGA	Gallus gallus (chicken) autosome
BC	Backcross
FIA	Flexible intercross analysis
SSc	Systemic sclerosis
AIL	Advanced intercross line
HWS	High weight selected line
LWS	Low weight selected line



## Introduction

Genes are functional units of DNA that an individual inherits from its parents. A gene determines parts of the observable characteristics (phenotypes) of an individual and specifies the composition of a protein. Each gene has a certain position on the chromosome called a locus. In most organisms, chromosomes occur in pairs (one coming from each parent) so there are two versions of a gene, which are called alleles. An individual carrying two identical alleles at a locus is said to be homozygous and an individual carrying two different alleles is heterozygous.

Some of the differences between individuals can be explained by differences in their genes. Related individuals tend to be more similar to each other than unrelated individuals because they share copies of the same alleles. The pair of alleles is called the genotype. It is generally accepted that the inherited genotype, transmitted epigenetic factors<sup>1</sup>, and environmental variation all contribute to the phenotype of an individual.

The basic principles of inheritance were established by Gregor Mendel. With his famous experiments on garden peas from 1856-1863, he formulated the fundamental rules concerning the transmission of hereditary characters that are known as the Mendelian laws of inheritance (Mendel, 1865). In

---

<sup>1</sup> Changes in gene expression caused by mechanisms other than changes in the DNA sequence. One example is DNA methylation.

1900, three scientists rediscovered the work of Mendel independently of each other (von Tschermak, 1951).

The traits Mendel studied were pea colour and texture, which are examples of qualitative traits. Qualitative traits are discrete and can have only a few different values. Initial successes were made in finding genes explaining qualitative traits because there were simple relationships between genotype and phenotype. Complex traits like body weight are quantitative, meaning that they show a continuous variation and it is believed that the variation is due to DNA variation at multiple genes or loci as well as the environment. Such loci are called quantitative trait loci (QTL). It is harder to find genes explaining a complex trait than a qualitative trait because each gene only accounts for part of the phenotypic variation and the rest is due to non-genetic factors (called environmental factors) as well as other genes (Falconer and Mackay, 1996). Fisher, Haldane, and Wright established much of the theoretical basis concerning the inheritance of quantitative traits.

A big challenge in genetics is to identify (map) the genes responsible for variation in complex traits. Studying how selective changes in allele frequency affect organisms can give us an increased understanding about the genetic regulation of diseases and metabolic traits. Increasing our knowledge about these traits can both help us to develop new treatments and diagnostics as well as to breed for increased food production.

Thousands of years of selective breeding of domestic animals has created a diversity of phenotypes among breeds. Domestic animals are therefore a unique resource for understanding the genetic basis of phenotypic variation (Andersson and Georges, 2004). Animals can be used as a tool to study human diseases and other complex traits since they provide an excellent opportunity to study traits in a controlled environment (Simmons, 2008). Human patients usually have different environmental backgrounds, which could have a large influence on the disease phenotype. Genetic studies in animal models also provide greater power than human studies thanks to the possibility to make experimental crosses because we

can choose which individuals to mate to get the most possible information. Furthermore, identified genes can be followed up with experimental validation and explorations of the molecular mechanisms. However, one big drawback with animal models of human diseases is that although the disease might resemble the human condition it could have a different cause, for example involve disruption of different pathways.



## Background

This thesis investigates the genetic regulation of metabolic and autoimmune traits in chickens using different outbred chicken models. New QTL mapping methods were applied and evaluated to analyse these complex traits.

First, two lines of chickens divergently selected for body weight at 56 days of age were used to study metabolic traits by generating and analysing different experimental crosses from these two lines (papers I-III). Second, a model for autoimmune avian systemic sclerosis was analysed (paper IV).

### Metabolic traits

Metabolic traits are related to, or involve aspects of metabolism. Since metabolism determines how efficiently food is converted into body mass these sorts of traits are of prime interest to farmers. Examples include growth rate, feed efficiency and fat deposition. An improved understanding of the genetic basis of these traits can both improve our knowledge of the genes that have been under selection and help identify genes and pathways to be included in future breeding programs.

### Autoimmune diseases

Autoimmune diseases occur when the immune system, which normally helps to protect the body from harmful substances,

becomes overactive and mistakenly recognises parts of the body as pathogenic (the immune system actually starts attacking parts of the body). This response results in persistent inflammation and damage to different tissues and organs of the body (Prokunina, 2004). Approximately 3% of the human population suffer from autoimmune disorders (Smith and Germolec, 1999) and women have a significantly higher risk of developing autoimmunity than men (Cutolo *et al.*, 1995). The mechanism of most autoimmune diseases is poorly understood but many autoimmune diseases have been associated with genetic, infectious and/or environmental factors. Some examples of autoimmune diseases are multiple sclerosis, Hashimoto's thyroiditis, systemic sclerosis and Type 1 diabetes mellitus (Hafler and Weiner, 1989; Hall and Stanbury, 1967; Kahalefp and Leroy, 1999; McCanlies *et al.*, 1998).

#### Systemic sclerosis

Scleroderma or systemic sclerosis (SSc) is a systemic autoimmune disease, meaning that the immune system attacks many different organs, tissues and cells of the body. SSc involves localised cell death (necrosis) leading to scarring (fibrosis) of skin and internal organs such as the gastrointestinal tract, lungs, kidneys and heart (Kuhn *et al.*, 2010). A clear visual phenotype of SSc in humans is fibrosis on the fingertips.

#### The chicken as an animal model

Since humans started to domesticate animals 9,000-12,000 years ago, animals used for breeding have been selected by their phenotype (Clutton-Brock, 1995). This approach is called selective breeding. By selecting for desired phenotypes, specific genotypes are enriched within that population, which makes animals useful as genetic models for studying the genes regulating these phenotypes. Selective breeding is also used in research. By selecting and breeding for desired traits or diseases, researchers can create an animal model showing the

selected trait. For example, if researchers would like to study a certain human disease and an apparently similar disease is seen in an animal population, they can select animals with the disease to increase the frequency of disease in the population. The selected group of animals, called a line, is then an animal model of the disease.

Since domestication from the Red jungle fowl (RJF), one of the ancestors of domestic chickens, a great phenotypic diversity has been developed within chickens making them interesting as genetic models (Fumihito *et al.*, 1994, Eriksson *et al.*, 2008). Selection has resulted in specialised egg layer and meat (broiler) chicken breeds. Selection has also resulted in a large variety of plumage colours, feather texture and size among breeds. Chickens are also useful for genetic studies because large numbers of offspring can rapidly be produced from each pair of parents, and phenotypic characteristics easily measured. Today there are many different experimental chicken lines and for example, there are three autoimmune chicken lines maintained in Uppsala: the OS-chicken line that gets autoimmune thyroiditis, the Smyth line chicken for vitiligo and the UCD-200 chicken line for systemic sclerosis (Dietrich *et al.*, 1999; Smyth, 1989; Gershwin *et al.*, 1981).

In this thesis, I have analysed systemic sclerosis in chickens, using a cross between UCD-200 and RJF chickens (paper IV). To study metabolic traits in chickens (paper I-III), I have used a chicken model of two lines selected for high and low body weight, respectively.

#### Chicken model for metabolic traits

Paul Siegel at the Virginia Polytechnic Institute and State University (USA) started a selection experiment back in 1957 where he selected white Plymouth rock chickens for high and low body weight, respectively, at 56 days of age. Since the experiment started, one generation is produced each year and the population is kept in sufficient numbers to minimize inbreeding (Liu *et al.*, 1994). The selection experiment has resulted in a notable selection response where the high weight

line (HWS) weighs almost twelve times more than the low-weight line (LWS) in the current generation (generation 54). Figure 1 shows a nine-fold weight difference between the Virginia lines at generation 40.



*Figure 1.* The photo shows a nine-fold weight difference between individuals from the high weight line (right) and low weight line at 56 days of age after 40 generations of selection. Today (generation 54), there is a twelve-fold weight difference between the lines (Photo: Dr. E.A. Dunnington).

The selection has also resulted in correlated responses in other traits including appetite, metabolic traits, reproductive traits and immune response (Siegel and Dunnington, 1987). Decreased body weight has resulted in later sexual maturity and lower egg production and fertility. Increased body weight has resulted in increased feed consumption, bone length, breast width, weight of fat and size of various organs and glands (Siegel and Dunnington, 1987).

Chicken model for human autoimmune systemic sclerosis

An abnormality in comb development in young White Leghorn chickens was discovered in 1942 (van de Water and Gershwin, 1985). Birds with the abnormality were selected to establish a small closed population, the University of California at Davis-line 200 (UCD-200). UCD-200 birds develop a hereditary SSc-



like disease, showing all the hallmarks of the human disorder, which makes this line a promising model to study genetic factors underlying the human disease. Studies performed in this population regarding the genetic, immunological and environmental influences on the disease have further confirmed clear similarities to the characteristics in humans (Beyer *et al.*, 2010, Gershwin *et al.*, 1981).

One to two weeks after hatching, UCD-200 birds have comb lesions at a current incidence of 97.5% (Beyer *et al.*, 2010). Disease progression starts with redness and swelling of the comb, leading to necrosis and in most cases, loss of the comb. These lesions are considered equivalent to fingertip necrosis in human SSc. Some of the animals also show decreased blood flow to toes and/or skin thickening at the neck. The early inflammatory stage subsequently progresses to a chronic stage characterized by fibrosis with excessive accumulation of collagen types I, III, and VI. Internally, the esophagus, small intestine, lungs and kidneys are affected, starting with redness and inflammation, followed by fibrosis (Wick *et al.*, 2006, Nguyen *et al.*, 2000).

## The chicken genome

Chickens have 39 pairs of chromosomes; one pair of sex chromosomes (ZZ or ZW) and 38 autosomes. The autosomes differ widely in size and are therefore divided into five macro, five intermediate and 28 microchromosomes. The smallest microchromosome is only ~2 Mbp while the largest macrochromosome is ~200 Mbp (Schmid *et al.*, 2005). It is still unclear why some of the avian genomes show this variation in size of the chromosomes. In chickens, females are the heterogametic sex<sup>2</sup> (ZW) while males are homogametic (ZZ). The avian sex chromosomes are not homologues to the X and

---

<sup>2</sup> Refers to the sex of a species in which the sex chromosomes are not the same.

Y chromosomes in mammals and evolved from a different autosomal chromosome pair (Fridolfsson *et al.*, 1998).

The chicken genome is relatively small, 1200 Mbp, which is only around 30-40% of the mammalian genome. In 2004, a high quality sequence of the chicken was released (International chicken genome sequencing consortium, 2004). Mammals and birds shared a common ancestor ~310 million years ago (Hedges, 2002), which makes the chicken sequence useful for increasing our understanding of the human genome through comparative genomics (Burt, 2005). More than 400 ultra-conserved regions greater than 200 bp sharing at least 95% sequence identity between chicken and human have been identified from comparative genomics (Burt, 2005). The chicken sequence is also useful for increasing our knowledge about the evolution and conservation of genes as well as for agricultural production.

## Genetic linkage maps

This section briefly summarises the theoretical background to the methods used to construct a genetic linkage map, i.e. to find the linear order and the relative recombination distances between loci. Detection and localisation of QTL on the genetic map is based on co-segregation between alleles at marker loci and the QTL.

### Genetic markers

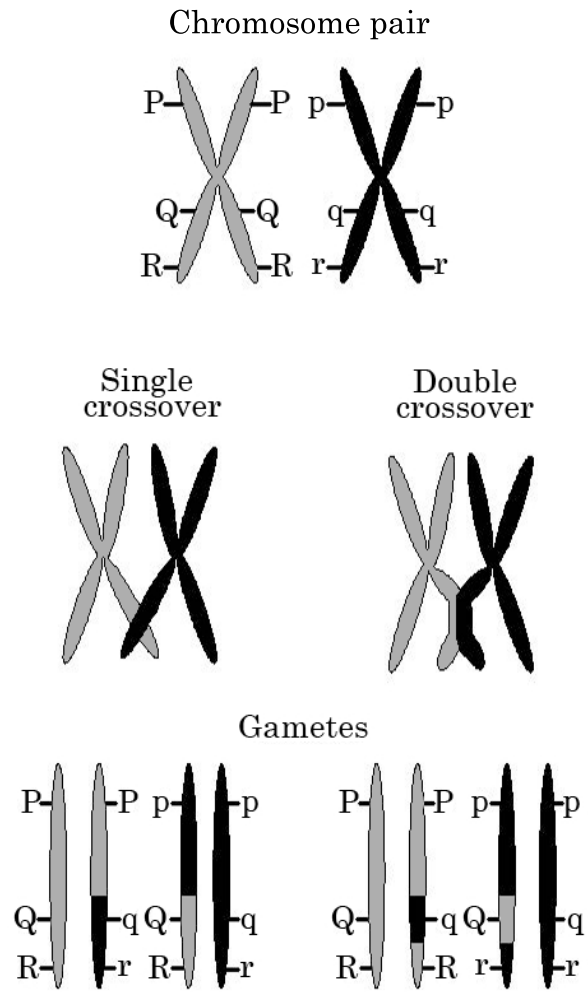
A genetic marker is a DNA variation (polymorphism) in the genome that can be observed. Early markers were measured by detectable phenotypes e.g. colour, or protein polymorphisms (e.g. blood type) but with technological improvements DNA polymorphisms themselves can now be measured as genetic markers. Microsatellites and single nucleotide polymorphisms (SNPs) are the most commonly used genetic markers. Microsatellites are repeating sequences in DNA, which are scored by the number of repeats, while a SNP is variation at a single position.

## Genetic linkage

In diploid species, adult organisms carry two copies of a gene (one allele on each chromosome) in each cell, with the exception of the germ cells (cells from which offspring are produced; gametes), which carry only one copy of each gene. During the production of gametes, cells divide according to a biological pathway called meiosis. Meiosis produces four gametes that each carry a random set of one copy of each chromosome.

Thomas Morgan Hunt (1866-1945) first described the idea of genetic linkage and hypothesised the phenomenon of crossing over. In the first stage of meiosis, both chromosomes in a pair are duplicated to produce two identical copies called chromatids. The four chromatids can physically exchange segments, which is called crossing over. The point where chromatids exchange genetic material is called a chiasma. Crossovers generate new combinations of alleles on a chromosome, which creates individuals with unique genomes. This process is called recombination (Figure 2).

Genetic linkage is the tendency of certain alleles at different positions to be passed together from one generation to the next. The degree of linkage between two loci depends on the frequency of crossovers that occur between them during meiosis. The closer loci are to each other, the less likely is it that a crossover between them will occur. Linked loci are said to be in linkage disequilibrium (LD) and form linkage groups. For loci in LD, combinations of alleles at the different loci (called haplotypes) occur more or less frequently in a population than would be expected by chance; that is, there is a non-random association of alleles at the loci. Linkage equilibrium is the converse of LD and describes the situation in which the frequency of haplotypes in a population is as expected if the alleles at each locus were combined at random.



*Figure 2.* Illustration of crossover and recombination. During the early stages of cell division in meiosis, chromatids from homologous chromosomes may exchange segments as shown above. This recombination produces genetic variation in germ cells.

Using recombination to construct a genetic linkage map

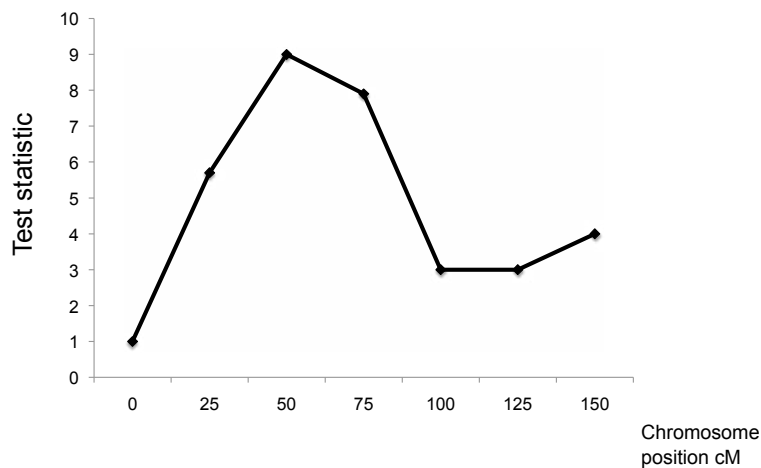
As mentioned earlier, a genetic linkage map shows the linear order of genetic markers along a chromosome and the relative distances between them based on linkage. To determine if two markers are physically linked together on the same chromosome, the recombination frequency between them is measured. The further away two markers are on a chromosome, the higher is the chance for recombination between them, which leads to lower LD. The recombination frequency is therefore used to measure the distance between markers. Map units are centiMorgans (cM), after Thomas Morgan Hunt. One cM is equal to an average of one recombination event in 100 meioses. Markers on different chromosomes are by definition non-linked and the frequency of recombination between them is 50%.

### Introduction to QTL mapping

Identifying genes affecting complex traits is a challenge and therefore a starting approach is to search for genomic regions harbouring genes affecting a trait rather than a specific gene. A chromosomal region affecting a quantitative trait is called a quantitative trait locus (QTL). QTL mapping is a useful tool for dissecting the genetic variance of complex traits and in particular it is used to locate and estimate the effect of a chromosomal region containing one or more genes affecting a quantitative trait. The detection of QTL relies on LD between the QTL and one or several genetic markers, which depends both on the information available from a set of polymorphic markers spread over the genome and on using populations with strong LD. Many candidate genes for human diseases, including asthma, diabetes and gallstones have been identified from QTL studies in mammalian disease models (Symula *et al.*, 1999; Podolin *et al.*, 2000; Bouchard *et al.*, 2000).

The general principle of QTL mapping consists of measuring the phenotype of interest for each individual in a population e.g. an experimental cross between two divergent lines (see

later) and collecting genotype information across the genome for the same individuals. A statistical association (high correlation) between variation in the measured trait and variation in the genotype (estimated based on information from the markers) is typically tested at regular intervals along the genome. A sufficiently high value of the test statistic indicates the presence of a QTL affecting the measured trait (Figure 3). The power to detect and map QTL depends on what fraction of the phenotypic variation is explained by a given locus and the size of the segregating population (Andersson and Georges, 2004).



*Figure 3.* Chromosome scan for QTL. The test statistic is estimated at each tested position along the chromosome. A high association between phenotype and genotype indicates the presence of a QTL.

The phenotypic variability of complex traits is affected by genetic as well as environmental factors. Phenotypic variance ( $V_P$ ) can be statistically partitioned into genetic ( $V_G$ ) and environmental variance ( $V_E$ ) components as follows:

$$V_P = V_G + V_E. \quad (1)$$

Genetic variance  $V_G$  can be subdivided into parts that result from the additive effects of alleles ( $V_A$ ), the dominance effects of interactions between alleles within a locus ( $V_D$ ) and the effects of interactions between alleles at different loci ( $V_I$ ) as

$$V_G = V_A + V_D + V_I, \quad (2)$$

(Lynch and Walsh, 1998). A QTL can be modelled by an allele substitution effect. The substitution effect is the average effect that is obtained by replacing one low effect allele ( $q$ ) for the high effect allele ( $Q$ ). When a gene is modelled as additive, the heterozygote is assumed to have a phenotype halfway between the two homozygotes and the combined effects of alleles at different loci are assumed to be equal to the sum of their individual effects. Dominance and epistasis measure deviations from these assumptions. Dominance occurs when the heterozygote ( $qQ$ ) differs from the mean of the two homozygotes, and epistasis is when the effect of a gene is dependent on the genetic background.

### Experimental crosses in QTL mapping

Whilst LD extends over long genomic segments in populations with closely related individuals, the range of LD is usually short in general populations. The shorter LD, the more markers you need to have a good chance of a QTL being in LD with one of them (Ke *et al.*, 2004). To increase the level of LD, experimental populations are often generated. A population created by crossing inbred lines is highly efficient for detecting QTL since the two lines have a high degree of homozygosity at marker loci and QTL, and their offspring will have alleles in high LD. The difference between parental lines of a mapping population must be sufficient for the trait of interest, both at the DNA and phenotype level, while the variation within each parental line should be low. If the two lines differ at the

phenotypic level for the trait, the probability that they differ genetically is high, although environmental effects can create large phenotypic differences without any genetic basis. Parents with similar phenotypes might also have genetic variation, as different genes could result in the same phenotype.

The more DNA variation there is between lines, the easier it is to find and select informative markers, i.e. markers that exhibit substantially different allele frequencies between the two lines. By crossing two lines with large differences in phenotype and genotype, where the variance within line is small, we can generate considerable genotypic and phenotypic variance in the offspring. Several types of experimental populations can be derived from crossing divergent lines and the ones used in this thesis will be presented in the following sections.

#### QTL mapping using outbred crosses

Experimental crosses based on inbred lines will have maximum LD between alleles at loci on the same chromosome. This makes these crosses highly efficient for QTL mapping. Most farm animals are, however, outbred. The major difference between using inbred-line crosses and outbred-line crosses for QTL analysis is that parents for the inbred lines are genetically uniform while outbred parents are genetically variable. Consequently, using crosses between outbred populations creates new challenges. Firstly, the parental lines could have the same alleles at certain regions. At these places markers will not always be fully informative (a fully informative marker has different alleles in the two lines that allow us to trace the inheritance from parents to offspring). Secondly, the same QTL allele may be in LD with different marker alleles in different parents and therefore inheritance must be traced separately for each parent (Lynch and Walsh, 1998). Thirdly, alternate QTL alleles may not be fixed in the parental ( $F_0$ ) lines. If this is the case, the power to detect QTL will decrease.

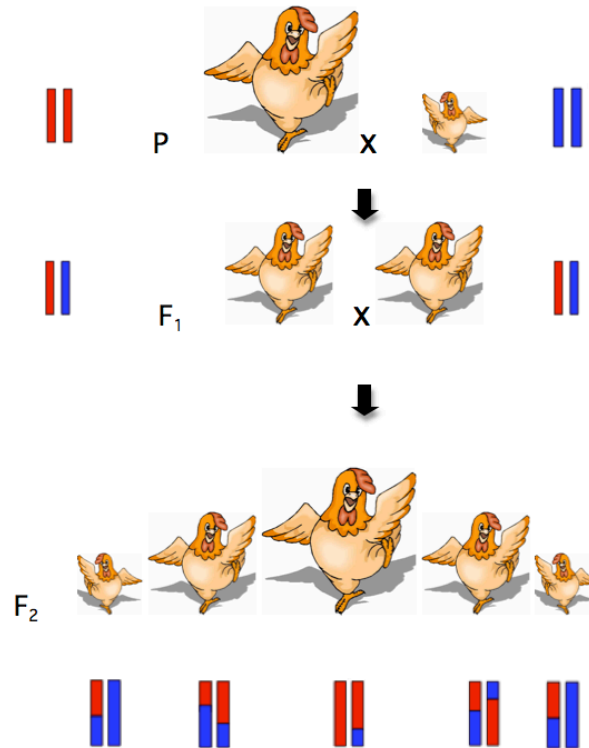


## F<sub>2</sub> and backcross design

In an F<sub>2</sub> design, parents from two lines are crossed to generate a first generation (F<sub>1</sub>). For inbred lines, the F<sub>1</sub> generation will be heterozygous at all loci that differ between the lines and have maximum LD. This F<sub>1</sub> population is then intercrossed with each other, which generates recombination and will break up LD. The resulting F<sub>2</sub> generation will differ phenotypically depending on which allele the individuals inherit at different loci. As a simple example, you could compare body weights in an F<sub>2</sub> chicken population. If most small birds have alleles with a certain line origin at a locus, while large birds have alleles from the other line at the same locus, there is an association between genotype and phenotype at this position (Figure 4). If this association is significant, you have found a QTL region that possibly contains genes affecting body weight. An F<sub>2</sub> design is a powerful approach for detecting QTL, but will usually give wide regions because the F<sub>2</sub> population have only undergone one generation of recombination and will have strong LD between markers. With strong LD between markers, the genotype information given by one marker will be correlated to other markers along the chromosome and the density of markers therefore does not need to be higher than one marker per 10 or 20 centiMorgans (cM) to be able to detect a QTL (Jensen, 1989).

For a backcross (BC) design, the F<sub>1</sub> generation is backcrossed to one of the parental lines, called the recipient line. For example, if you start with a cross between two divergent lines, one showing a high incidence of a certain disease and another that is healthy, you would backcross the F<sub>1</sub> generation to the diseased line to increase disease frequency in the offspring generation. An F<sub>2</sub> design is generally very powerful because both F<sub>1</sub> parents are expected to segregate at the trait loci. However, if the disorder is dependent on interactions between several loci and the incidence of disease is low in an F<sub>2</sub> population, a BC design can be more powerful if it generates more affected individuals. For example, if a trait is caused by recessive alleles at four

unlinked loci, in an  $F_2$  design  $(1/4)^4 = 1/256$  of the progeny would be diseased, whereas  $(1/2)^4 = 1/16$  of the progeny would be affected in a BC design (if backcrossing to the diseased line).



*Figure 4.* Schematic overview of an  $F_2$  intercross between two divergent chicken lines. Parentals (P) from the genetically divergent lines are intercrossed to generate an  $F_1$  population, which is further intercrossed to produce recombinant offspring ( $F_2$  population). The bars represent a pair of chromosomes and the colours show from which parental line each region is inherited.

Fine mapping using an advanced intercross line

A second challenge with complex traits, after detecting QTL, is to identify the genes that underlie them. One problem is that it is hard to detect the exact position for QTL as this is usually a relatively large region when the LD in experimental crosses is high. The broader the QTL region, the higher is the chance that many genes are in that region. This makes it is hard to find which gene, or genes, are causing the phenotype. Narrowing down the QTL region is thus an important step towards the identification of causative genes. One way to narrow down the regions is to continue generating additional generations in an experimental cross by repeated intercrossing, e.g. using  $F_2$  individuals to generate an  $F_3$  generation and so on, to form an Advanced Intercross Line (AIL) (Darvasi and Soller, 1995). The major difference between the  $F_2$  and the following generations of the AIL is the increase in recombination events that decreases LD between markers and QTL. Since an AIL will have more recombination events, markers will be less correlated and a higher density of markers will be needed to detect QTL.

In order to fine map and reanalyse previously found QTL regions in the HWS and LWS cross (paper I), an AIL population was generated and analysed in paper II.

Introgression lines

QTL mapping in intercrosses between divergent lines is a commonly used strategy to identify genetic loci that regulate complex traits. The primary advantage of this approach is that it is a powerful way to detect the loci determining the difference between the parental lines. Disadvantages are that the loci are usually mapped with low resolution due to the limited amount of recombination in the population and that the estimated genetic effects of the loci are estimated only in the intercross genetic background. To study the genetic effects of identified loci in a more homogenous background it is sometimes desirable to transfer important genes or QTL from a donor line into the genetic background of a recipient genotype. This can be done using recurrent backcrossing of

the donor line to the recipient line to generate a so called introgression line. Introgression is initiated by generating an  $F_1$  population from the two founder lines. The  $F_1$  individuals are thereafter backcrossed for a number of generations to the recipient line. At each generation, genetic markers within the QTL region are used to identify the individuals that carry donor alleles within the QTL to be used for breeding the next generation. The final step in the introgression phase is to generate homozygote individuals by intercrossing the last backcross generation (Gama *et al.*, 1992; Groen and Smith, 1995).

In paper III, we generated an introgression line between the HWS and LWS lines. The process involved introgressing QTL from the LWS line (donor) into the HWS line background (recipient).

### QTL mapping methods used in this thesis

Three QTL mapping methods were used to search for QTL affecting avian systemic sclerosis in paper IV. Two methods that are designed for analysing outbred crosses, a linear regression model (Haley *et al.*, 1994) and a variance component method (Rönnegård *et al.*, 2008), and one method accounting for epistatic interactions. The variance component approach was also used in paper II. These methods are further described in this section.

#### QTL mapping by linear regression

As described in Figure 4, the founder parents ( $F_0$  generation) are expected to have large genetic variance between lines and small genetic variance within lines. When crossing them we will generate an  $F_1$  generation that will inherit one of each chromosome pair from each parent. The second generation ( $F_2$ ), will show recombination between the parental chromosomes, one region of a chromosome can come from one line and another region from the other line. Different chromosome regions will inherit either of the two possible line

origins and we assess which line origin is inherited at each position along the chromosome with genetic markers. In an  $F_2$  population there can be three genotypes at each locus. These can be represented as AA (homozygote for line A), BB (homozygote for line B) and AB for the heterozygotes. In a backcross, the  $F_1$  generation is backcrossed to one of the parental lines, called the recipient line. A BC generation will therefore only have two genotypes: homozygote AA for the recipient line A, or heterozygote AB. The genotype will vary between individuals and between loci, meaning that the relationship between phenotype and genotype will vary between positions and we can search for the position that best explains the phenotypic variation (Haley and Knott, 1992). The genetic substitution effect is modelled as a fixed effect with the following equation:

$$y = X\beta + e, \quad (3)$$

where  $y$  is the vector of phenotypes,  $\beta$  is the vector of fixed effects,  $X$  is the incidence matrix for the fixed effects (genetic effects + other possible fixed effects such as sex) and  $e$  represents the residuals.

As described previously, the model is fitted at each tested chromosome position. The significance thresholds for the scan are derived by means of permutation, which is a way of empirically determining significance when the distribution of the test statistic is not known. For a permutation test, the trait values are repeatedly randomly rearranged relative to the genotypes. After each shuffle, the data is analysed for QTL effects and the maximum value of the test statistic is recorded. These values are ordered to form an empirical distribution for the test statistic under the null hypothesis from which the critical value can be obtained. For example, if we chose an overall significance level of 0.05 and run 1000 permutations, the 950th highest value from the permutations will be our estimate of the threshold (Churchill and Doerge, 1994). 5% (0.05) is a conventionally accepted significance

level; if the deviation is so large that the probability of it occurring by chance is less than 5%, we reject the null hypothesis and conclude that there is a significant effect.

Variance component QTL mapping

QTL mapping by linear regression is designed for experimental populations generated by line crosses and relies on the assumption that the lines are fixed for alternative QTL alleles. If this condition is fulfilled, the linear regression approach is powerful but if the QTL alleles are not fixed, it may be more powerful to use an approach that accounts for this.

A variance component approach is more flexible for detecting QTL in various types of populations and will account for possible segregation within founder lines. QTL effects are modelled as random effects (Fernando and Grossman, 1989; Goldgar, 1990) rather than fixed effects as in the regression method. A mixed model is fitted since fixed effects like sex are usually included in the model as

$$y = X\beta + Zv^* + e, \quad (4)$$

where  $y$  is the vector of phenotypes,  $\beta$  is the vector of fixed effects,  $X$  is the incidence matrix for the fixed effects,  $Z$  is an incidence matrix of size  $n \times m$  relating individuals to the QTL alleles in the base generation,  $v^*$  is the vector of  $m$  normally distributed base generation QTL alleles with variance  $\frac{1}{2} \sigma^2$ , and  $e$  represents the residuals.

The most likely phases of the marker genotypes are estimated at each tested genomic location using a haplotype-based method (Besnier *et al.*, 2009) from which a genotype IBD matrix is constructed. An IBD matrix ( $\Pi$ ) is an  $n \times n$  matrix that contains the expected number of alleles IBD between all pairs of the  $n$  individuals in a given population (Lynch and Walsh, 1998). When used in a mixed linear model with a random QTL effect, the IBD matrix is proportional to the covariance matrix of the QTL effect. Haplotypes are used to

calculate IBD values if these are robust and if not, individual marker genotypes are used instead (Rönnegård *et al.*, 2008).

As in regression based QTL mapping, the model is fitted at each tested position in the genome. The likelihood function follows a multivariate normal distribution.

#### Epistatic QTL mapping

Most QTL mapping experiments are based on assumptions of additivity, i.e. that the effects of genes are independent of one another and are summed together to produce the phenotype of an individual. However, since the late 1970s, biologists started to rediscover and explore the importance of interactions among loci (epistasis) or non-additive genetics, i.e. that the influence of genes on the phenotype depends upon the context of other genes (Wolf *et al.*, 2000). If a locus affecting a specific trait does not cause the same trait phenotype in all individuals, one possible reason for this could be that the effect of one gene may depend upon alleles at other loci. By estimating the consequences of both significant individual and epistatic effects it may be possible to better explain the total phenotypic variation (Carlborg and Haley, 2004). These genetic interactions are often ignored in QTL mapping but several studies have showed that the power to detect QTL is decreased if epistasis is ignored (Fijneman *et al.*, 1996, Long *et al.*, 1996, Li *et al.*, 1997, Carlborg, 2002). The power to detect QTL can be increased by searching for multiple QTL and taking into account both the individual QTL effects and the effect of their interactions (Carlborg, 2002).

For complex traits, epistasis describes any interaction between two or more loci. If a trait is regulated by epistatic interactions, the phenotype of any genotype cannot be predicted by summing the effects of individual loci. Carlborg *et al.*, (2006) found four QTL with significant epistatic effects in an F<sub>2</sub> cross between the previously mentioned HWS and LWS chicken lines. This four-locus network accounts for 45% of the body weight difference between these lines.

In this thesis I have further investigated the effect of this four-locus network, first on other metabolic traits in the same  $F_2$  cross (paper I), then by fine mapping the network in an AIL (paper II) and generating an introgression line including three of these four QTL. I have also included a model accounting for epistatic interactions when mapping genes for avian systemic sclerosis in a chicken cross (paper IV).



## Aims of the thesis

*The objectives of the thesis were to*

- Investigate how an interacting network of four QTL that have been found to affect body weight in chicken regulates other metabolic traits, specifically to
  - include more markers and confirm the results from an earlier study,
  - analyse the affect of the four QTL network on other metabolic traits,
  - analyse data from a chicken AIL population to replicate and fine map the four-QTL network,
  - introgress three of the QTL into a homogenous background to better understand their effect on metabolic traits.
  
- Map QTL affecting autoimmune diseases in chickens. In particular to
  - apply and evaluate new statistical methods to analyse an experimental chicken model for human systemic sclerosis.



## Summary of investigations

### Analysing metabolic traits in chickens (papers I-III)

Body weight and other metabolic traits, e.g. breast muscle weight and abdominal fat, were studied in two outbred chicken lines divergently selected for body weight at 56 days of age (papers I, II and III). A study by Jacobsson *et al.* (2005) found 13 QTL affecting body weight in an intercross between these lines and Carlborg *et al.* (2006) found that a network of four interacting loci explains nearly half of the weight difference between the lines at 56 days of age. The central locus in this network is located on chicken chromosome 7 (*Growth9*) and acts by releasing the genetic effects of loci located on chromosome 3 (*Growth4*), chromosome 4 (*Growth6*) and chromosome 20 (*Growth12*).

### Paper I. Genetic analysis of metabolic traits in an intercross between body weight selected chickens

The original study of the four-QTL network reported effects on body weight and fat deposition, but did not investigate the effects on other metabolic traits (Carlborg *et al.*, 2006). This study was designed to explore how the network of four QTL regulate other metabolic traits including breast muscle, abdominal fat and shank weight, rather than searching for novel loci associated with these traits. The aim was to see how these four loci influence traits that might increase or decrease

body weight and also to refine the original study by including a larger number of genetic markers in the QTL regions.

#### Methods

795 F<sub>2</sub> individuals were genotyped for 495 markers covering ~3100 cM in 25 linkage groups. Descriptive statistics for the traits analysed are summarised in Table 1.

Table 1. Summary of phenotypic data for F<sub>2</sub> intercross of HWS and LWS at generation 40. Values are mean mean±standard deviation for n chickens (Jacobsson *et al.*, 2005, Park *et al.*, 2006).

Trait	n	Weight [g] ± SD
Body Weight [g]		
Hatch	795	27.8 ± 2.1
14 days	795	75.2 ± 14.9
28 days	795	179.1± 56.8
42 days	795	365.5±113.1
56 days	795	621.7 ± 186.9
70 days	795	943.3 ± 262.2
Body traits at 70 days of age [g]		
Abdominal fat	402	5.4 ± 4.1
Shanks	405	42.5 ± 11.9
Breast <sup>1</sup>	201	91.1 ± 28.8
Lungs	405	6.5 ± 2.2
Spleen	401	1.4 ± 0.5
Bursa	405	1.9 ± 0.7

<sup>1</sup> Breast weight includes pectoralis major + pectoralis minor.

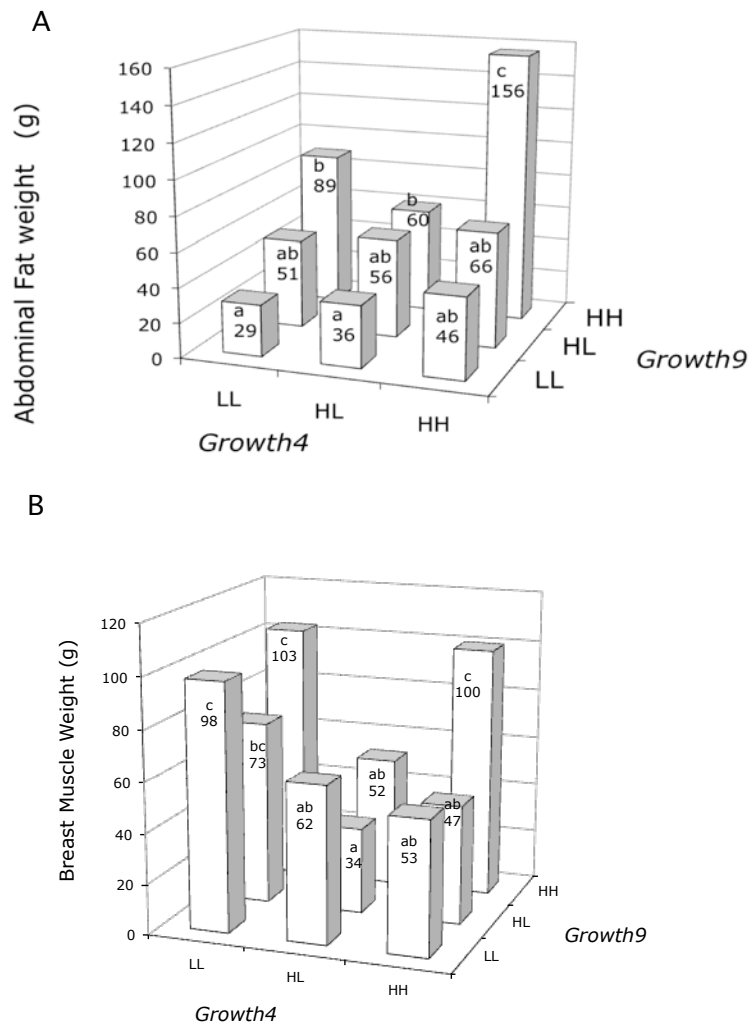
The original study (Carlborg *et al.*, 2006) showed that the effects of the loci on chromosome 3, chromosome 4 and chromosome 20 are dependent on the genotype on chromosome 7. Multiple regressions were therefore used to model the effects of the loci on chromosome 3, chromosome 4 and chromosome 20, with separate models for each genotype on chromosome 7 (i.e. homozygote high line (HH), homozygote low line (LL) or heterozygote (HL)). The following model was used:

$$Y = \mu + Z\beta + \beta_1 X_{Growth4} + \beta_2 X_{Growth6} + \beta_3 X_{Growth12} + e, \quad (5)$$

where  $Y$  is the phenotype for each genotype group, i.e. HH, LL or HL for the locus on chromosome 7 (*Growth9*),  $\mu$  is the mean of the analysed group,  $\beta$  is the vector of fixed effects of sex,  $Z$  is the incidence matrix for the fixed effects,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are the additive effects,  $X_{Growth4}$ ,  $X_{Growth6}$  and  $X_{Growth12}$  are indicator variables for the homozygote genotype for *Growth4*, *Growth6* and *Growth12*, respectively, and  $e$  is the error term.

#### Results and discussion

The study confirmed the findings of the four-locus network and our explorations of the effects of the network on the measured traits indicate that the network mainly increases the body weight through increased growth of all the measured organs rather than by changing the relative size of different organs. A significant two-locus interaction between the loci *Growth9* and *Growth4*, in addition to contributing to the general increase in body weight, also alters the relative amount of breast muscle and abdominal (Figure 5).



*Figure 5. Two-locus genotype-phenotype maps for (A) abdominal fat (g) and (B) breast muscle weight (g): both phenotypes were corrected for body weight at slaughter. LL, low weight selected (LWS) homozygote; HH, high weight selected (HWS) homozygote; HL; heterozygote. Different letters indicate genotype combinations that are significantly different from each other ( $P < 0.05$ ).*

The QTL regions were too large to reveal specific candidate genes. However, some genes known to affect body weight are located in the regions. The gene encoding *latent transforming growth factor- $\beta$  binding protein 1* (*LTBP1*) is located in the QTL on chromosome 3 and *insulin-like growth factor binding proteins 2 and 5* (*IGFBP2* and *IGFBP5*) are located in the QTL on chromosome 7 (Öklu *et al.*, 1998; Wolf *et al.*, 2000).

## Paper II. Fine mapping and replication of QTL in an outbred advanced intercross line

Outbred crosses of divergent selection lines are a common experimental design for QTL mapping. Both the  $F_2$  and the BC population designs are powerful strategies to identify QTL, but will normally find regions so large that it is hard to identify candidate genes. In these designs, linkage disequilibrium is strong along the chromosomes, since only one generation of recombination occurs (gametes from the  $F_1$  population). The strong LD between markers makes it possible to detect QTL by linkage even with low marker density.

By fine mapping, it is possible to narrow down QTL to smaller regions, helping to identify likely candidate genes. In paper II, we generated, genotyped and analysed a nine-generation chicken AIL population to fine map and confirm nine of the QTL regions reported by Jacobsson *et al.*, (2005). The major difference between the  $F_2$  and the subsequent generations of the AIL is the increase in recombination events, which will decrease LD between markers.

### Methods

A subset of 304 segregating SNPs were selected from the nine QTL regions to discriminate between regions inherited from the high and low weight lines. To fully determine the line origin, a SNP should ideally be fixed for different alleles in the two lines, but can still be useful if one allele has a high frequency in one of the lines and a low frequency in the other line. This situation rarely occurred for the markers available for this

study and markers were therefore selected as follows: First, differences in marker allele-frequencies between the HWS and LWS were evaluated for all markers in the QTL regions. Secondly, markers were selected based on differences in marker allele frequencies between the lines and on their positions, aiming at an even distribution within the chromosome segments. All individuals in the AIL ( $n = 1529$ ) were genotyped for these markers using the Golden Gate assay (Illumina, CA, USA) at the SNP technology platform in Uppsala (Sweden).

To map QTL in this outbred AIL population, a variance component approach that accounts for correlation among QTL alleles within the founder lines was used. When detecting QTL by a variance component method, the covariance matrix of the random QTL effect is estimated by an IBD matrix, i.e, an  $n \times n$  matrix that, at a given genomic position, contains the expected number of alleles IBD between all pairs of individuals in the studied population.

#### Results and discussion

All of the nine QTL originally detected in the  $F_2$  population also showed effects in the AIL; five significantly confirmed and the remaining four showed suggestive evidence. All of the original QTL were confirmed as single loci, except for the QTL on chromosome 7 that split into two linked QTL. Several QTL peaks in the AIL are narrower than the original QTL illustrating the increase in resolution when using an AIL compared to an  $F_2$ . These results strengthen the original findings, and validate AIL as a valuable tool to explore the genetic basis of complex traits.

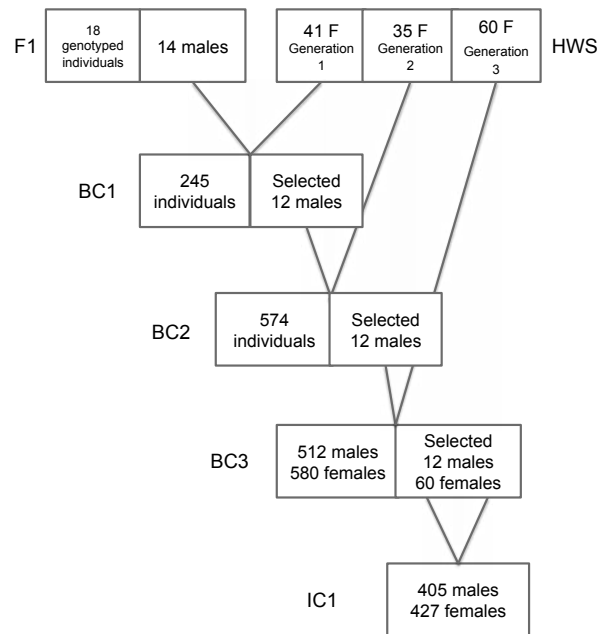


### Paper III. Using marker assisted introgression to replicate multiple epistatic loci with large effect on juvenile body weight in chicken

Once the gene network was replicated and a higher resolution of the QTL positions was obtained in the ALL, we designed a scheme to breed a QTL introgression line in order to facilitate studies of the genetic effects of this network on metabolic phenotypes in more detail. The process involved introgressing genomic segments (QTL) from the low body weight selected line for three of the four loci in the interaction network, into the high body weight selected line.

#### Methods

In Figure 6, we provide a detailed description of the breeding plan for the three-locus introgression line. In short, an  $F_1$  population was generated by crossing 14 LWS males and 56 HWS females from generation 50 of the Virginia lines (Dunnington and Siegel, 1996). To trace the inheritance of LWS alleles through the recurrent backcrossing to HWS females, a SNP marker panel of 32 markers was developed. 12 males from the  $F_1$  population, that were heterozygous for the three QTL regions selected for introgression, were selected to backcross to HWS females and produce the first backcross generation. In total three backcross generations (BC1-BC3) were generated with 4-5 batches per generation. In BC3, both males and females were genotyped for the markers in the QTL regions. Birds to use as breeders were selected based on their heterozygosity in the QTL regions and intercrossed (BC3xBC3) to generate the first intercross generation (IC1). The high line was used as the recipient line (background) since this line has a higher fertility and survival rate than the low line. To increase hatchability, semen was pooled with 3 males in each pool.



*Figure 6.* Breeding scheme for the introgression line, showing the number of individuals for each generation (F1-IC1), and the number of males and/or females (F) selected for breeding. In the parental generation (F1) we genotyped 18 individuals from which 14 randomly selected males were used as breeders to the next generation. Sex ratio is presented for BC3 and IC1.

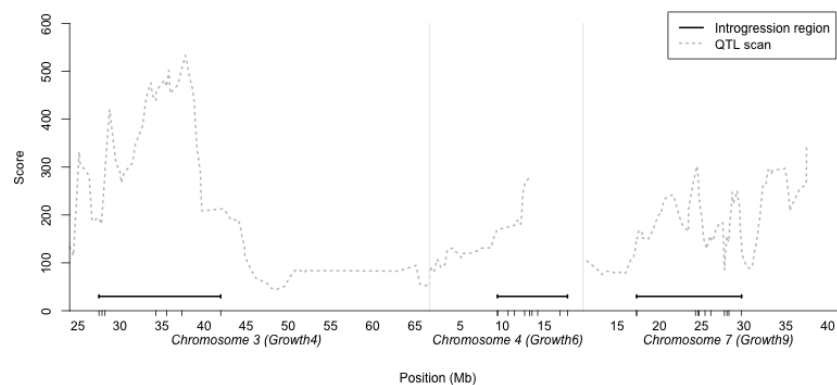
#### Results and discussion

Figure 7 shows the QTL profiles obtained in paper II and the segments that were introgressed. The figure also shows the location of the markers genotyped in the regions.

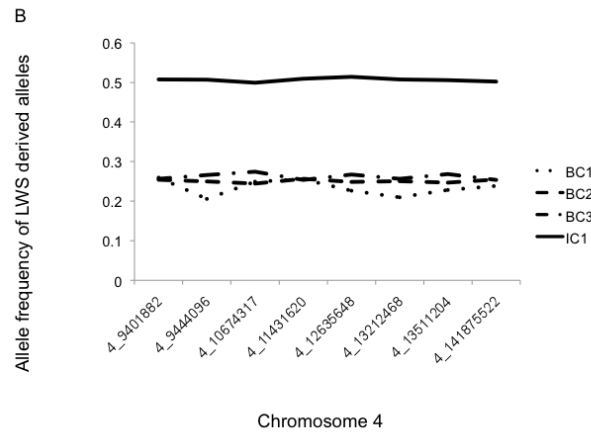
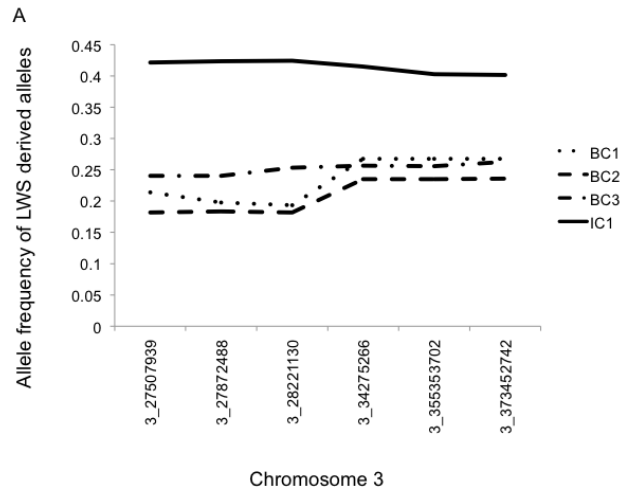
To study the efficiency of the introgression procedure across generations, frequency of the LWS allele was traced for all the 32 genotyped markers in all generations (BC1-IC1). Figure 8 illustrates that the LWS allele frequency, as expected in a successful introgression experiment, remains constant at

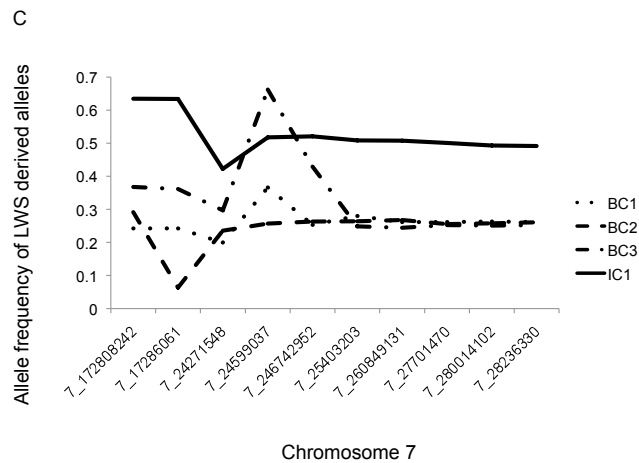
25% for the selected QTL regions in the BC1-BC3 generations. This is because half the individuals should be heterozygous in the region and the others should be homozygous HWS. Some of the first markers on chromosome 7 had low information content, which gave erratic frequency estimates. The IC1 is generated by intercrossing BC3 males and BC3 females selected for heterozygosity at the three QTL. As a result the allele frequency in this generation is expected to be near  $p=q=0.5$  for LWS and HWS alleles. When selecting BC3 breeders to generate IC1, there was a shortage of females that were heterozygous for the entire QTL segment in all three regions. It was therefore decided that a smaller number of females recombinant in the region on chromosome 3 would be used as parents for the IC1 generation. As a result, there is a lower frequency of LWS alleles (40-45%) in this region (Figure 8A).

In a near future, use of this confirmed introgression line should facilitate detailed studies of the effects of genetic interactions on complex traits.



*Figure 7.* Chromosomal segments on chicken chromosome 3, 4 and 7 selected for introgression in relation to the statistical support-curve for the corresponding QTL in these regions from a replication study in an nine generation Advanced Intercross Line. Lines under the x-axis indicate the location of the markers genotyped in the regions (Besnier *et al.*, 2011).





*Figure 8.* The LWS allele frequency across the four generations (BC1-IC1) in the chicken introgression line. Dotted lines show the backcross generations and the solid line IC1. A, B and C give the allele frequency at each marker in the introgressed regions on chromosome 3, 4, and 7, respectively. The scored markers in each segment are provided on the x-axis and are named according to the system “Chromosome”\_ “Position on chromosome in Mbp”.

#### Paper IV. Mapping systemic sclerosis genes in an outbred chicken intercross

In paper IV, I analysed an outbred backcross (BC) population generated from the UCD-200 chicken line, selected for an autoimmune disorder similar to systemic sclerosis in humans, and the red jungle fowl (RJF).

##### Methods

A BC design was used since an earlier  $F_2$  population with UCD-200 chickens showed a low incidence of affected offspring (4%), compared to a BC where approximately 50% of the offspring displayed the disease (Abplanalp *et al.*, 1990). We

analysed the BC population (n=471) to search for QTL affecting avian systemic sclerosis with three different statistical models, a generalised linear model, a variance component model and a two-way QTL interaction model. We measured avian autoimmune systemic sclerosis as a binary trait, sick or healthy. Systemic sclerosis is therefore not a quantitative trait, but it is reasonable to assume that there is underlying quantitative variation and by accounting for the binary structure it is statistically possible to use QTL mapping for analysing this trait as well.

#### *Linkage map*

A genetic linkage map was built using the CRI-MAP software (Green *et al.*, 1990). The map included 299 SNP markers evenly distributed over 26 autosomes and 27 SNP markers located on the Z chromosome.

#### *Generalised linear regression*

First, we analysed the data using a generalised linear regression model. Disease status was regressed on the probability that both alleles at a specific locus along the chromosome originated from UCD-200, calculated at 1 cM intervals across the area of the genome covered by the linkage map (Haley *et al.*, 1994). These probabilities were estimated using a method called triM (Crooks *et al.*, 2011), which uses a computationally efficient algorithm allowing information from many markers to be included.

As the trait was binary, a binomial error structure and a logit link function were used. Regressions were performed in R using the packages R/qtl (Broman *et al.*, 2003) and qtl.outbred (Nelson *et al.*, 2011). Qtl.outbred extends the functionality of R/qtl to handle data from outbred line crosses. Fixed effects of sex and batch were included in the model and allele substitution effects were fitted. Dominance effects could not be fitted because of the backcross design.

The logarithm of odds (LOD) score, which is related to the likelihood ratio test statistic, was used and significance

thresholds were established using randomisation testing with 1000 permutations (Churchill and Doerge, 1994). A 5% genome-wide threshold was used as significance threshold and a 20% genome-wide threshold for suggestive QTL. The QTL effects are given as the change in odds ratio of disease (OR) that is how much the probability of being diseased is increased relative to the probability of being healthy. ORs are typically used to present results from binary models.

#### *Variance component approach*

Secondly, the data was analysed using Flexible intercross analysis (FIA), which is a variance component based method that accounts for segregation within founder lines (Rönnegård *et al.*, 2008). We analysed the data with this approach since we were working with an outbred cross and wanted to account for potential segregation within lines. Significance testing in FIA is based on a score statistic to save computational time when deriving empirical significance thresholds using randomisation testing. FIA assumes that the phenotypes are normally distributed.

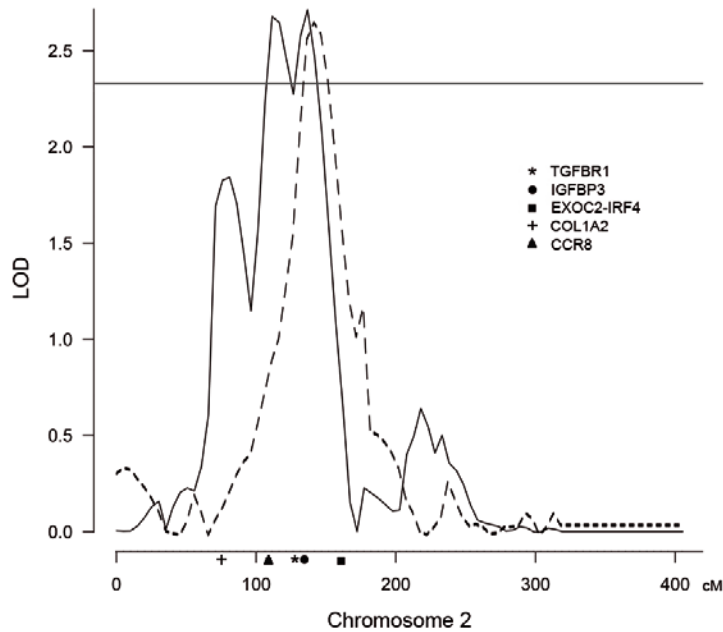
#### *Epistatic analysis*

A scan for pairwise epistasis was conducted using the *scantwo* function in R/qtl (Broman *et al.*, 2003), using the same line origin probabilities described earlier. We therefore treated the phenotype as being normally distributed. *Scantwo* evaluates every possible pair of genomic positions and calculates several LOD scores allowing for different tests depending on whether the positions were significant when tested individually. Significance thresholds for all scores were established by randomisation testing using 1000 permutations.

#### Results and discussion

This study identified a significant difference in disease frequency between the sexes, with the homogametic males being more affected. In humans, females, which are also the homogametic sex, are more affected. There is no sex

difference in pure UCD-200 populations, but in crosses including UCD-200, males are more affected. Interestingly, we observed less disease in females with a W chromosome inherited from the UCD-200 founder line, suggesting that a protective effect of the W chromosome might have evolved in UCD-200 chickens. Three suggestive autosomal susceptibility QTL were also mapped in the chicken genome. Interestingly, several orthologues<sup>3</sup> of human SSc candidate genes are located within two of these loci (Figure 9).



*Figure 9.* QTL-scan of chicken chromosome 2 using logistic regression. Solid line shows profile for early SSc and dashed line profile for late SSc. Estimated positions for the following orthologous genes are marked with symbols: *COL1A2* at 75 cM, *CCR8* at 108 cM, *IGFBP3* at 130 cM, *TGFBR1* at 133 cM and *EXOC2-IRF4* at 160 cM. The horizontal line shows the 20 % genome-wide suggestive threshold (2.3).

<sup>3</sup> Orthologues are defined as homologous segments that arose from a common ancestral gene during speciation in different species. Orthologues between species do not always have similar gene functions.



## Future research

### Ongoing work

#### Metabolic traits

Further studies are needed to explore the physiological mechanisms by which the discovered four-QTL network affects metabolic traits. Power studies done on the data used in paper I (not included in the article), showed that we need more individuals with the most informative genotypes (homozygote HWS and homozygote LWS) to have enough power to find effects on other metabolic traits. The introgression line (paper III) was developed for this reason and will be extensively phenotyped including measurements of conformational phenotypes such as body and tissue weights. In addition, tissues will be harvested to evaluate the expression profiles in various tissues in individuals with alternative genotypes at the introgressed QTL.

#### Autoimmune traits

There was a significant difference in disease frequency between the sexes in the autoimmune cross, with the homogametic males being more affected. In humans, females, which are also the homogametic sex, are more affected and understanding the reasons for this is an interesting subject for further study. The large effect of the W chromosome might

mask the effect of other QTL. Therefore, we will reanalyse the data fitting the W chromosome as a fixed effect with three classes: W from the UCD-200 line, W from RJF, and males without W. We will also fit a model for interaction effects between autosomal loci and the W chromosome, which will test whether these loci have different effects depending on the W chromosome origin.

### Near future

We can clearly see that avian systemic sclerosis is inherited and therefore regulated by genes. However, our study did not give any strong correlation between genotype and phenotype. The three QTL found in the autoimmune study were suggestively associated with disease and extending the study with more individuals might provide stronger evidence for these regions. Especially for complex traits, it is important to generate a large population as if many genes contribute to the trait each gene may explain only a small amount of the overall variance.

It might be valuable to construct different intercross populations. In our study, the UCD-200 chicken line was crossed with RJF chickens and further backcrossed to UCD-200, which (assuming QTL fixation in the lines) will only generate two genotypes at each locus, the homozygote state for UCD-200 and the heterozygote, compared to three genotypes generated in an  $F_2$  population. This backcross design prevents us from estimating potential dominance effects. If there are dominance effects explaining avian systemic sclerosis, it would be better to produce a large  $F_2$  population. The studied cross also consisted of many founders (8 RJF and 20 UCD-200). This might have made it harder to find genetic effects, if there were different QTL causing disease in different founders. If we generated a population with

fewer UCD founders, we might have a better chance to detect genetic effects.

Many microchromosomes are still uncovered in the systemic sclerosis study, and to gain further information, a complete genome scan that covers all chromosomes should be performed. There were very few markers on some chromosomes, which can especially reduce the power to find epistasis. When possible, the remaining chromosomes should be included since they may carry important genes regulating the disease, either individually or acting epistatically with other loci in the genome.

Another aspect that may have affected the power of our study is the accuracy of the phenotyping. Any errors in phenotyping a binary trait have a strong impact, resulting in a change in classification e.g. from healthy to diseased. In a quantitative trait, measurement errors will shift the value but the relative effect will be much smaller. If we could find a measurable intermediate quantitative trait, e.g. cytokine levels, we might improve power.

Access to dense maps should open up the possibility of new approaches for QTL detection. If the marker density is high enough, one could screen a number of individuals from different populations and look for regions with consistent differences indicating selective sweeps. In this case it would be better to compare UCD-200 with its founder breed, White leghorn, where the biggest differences between the lines should be related to avian systemic sclerosis.

### Looking ahead

It would be interesting to investigate possible genetic by environmental interactions on both these chicken models (Virginia lines and UCD-200) to further investigate the traits. We could keep crosses in different environments to investigate if they develop the same phenotypes and if the same QTL are found. This could be of considerable interest for autoimmunity since it is believed that the environment has a substantial

effect on disease. Exposure to different antigens can give large differences in the disease phenotype. Genetically predisposed individuals might not develop disease if the environment is very clean and would then be missed in phenotyping. Environmental factors like temperature, food and diseases can have also have a large effect on body weight. The design would be to generate one cross and then split the  $F_1$ s into different environments for generating the  $F_2$ . Finding the same QTL in both environments will give greater support for their effects. Comparing the two populations will give us the opportunity to study gene by environment interaction.

## Conclusions

### Paper I

Our explorations of the effects of the four-QTL network on other metabolic traits indicate that the major contribution of the network is to increase body size through an overall increase in growth of all organs rather than by changing the relative amount of different body components. An exception is the interaction between the loci on chromosome 3 and 7 that in addition to its contribution to the general increase in body weight, also alters the relative amount of breast muscle and abdominal fat.

### Paper II

An AIL was genotyped and analysed to fine map the four QTL studied in paper I. All of the QTL originally detected in the  $F_2$  population were confirmed. Confirming the QTL and fine-mapping their location using an AIL strengthens the original findings, and validates AIL as a valuable tool for exploring the genetic basis of complex traits.

### Paper III

Once the QTL network studied in paper I was confirmed in the AIL population (paper II) we designed a new strategy to study the network in more detail. The first study (paper I) showed

that we need more individuals with the most informative genotypes, which were HWS homozygotes and LWS homozygotes, to be able to detect any effects on other metabolic traits. To study the genetic effects of identified loci in a more homogenous background we generated an introgression line where we transferred three of the QTL from the LWS line into the genetic background of the HWS genotype. The introgression worked as planned and will be extensively phenotyped including measurements of conformational phenotypes such as body and tissue weights.

#### Paper IV

Here we illustrated the usefulness of the UCD-200 chicken as a model for human autoimmune systemic sclerosis. We found a significant difference in disease frequency between the sexes, with the homogametic males being more affected. In humans, females, which are also the homogametic sex, are more affected and understanding the reasons for this is an interesting subject for further study. Interestingly, we observed less disease in females with a W chromosome inherited from the UCD-200 founder line. This indicates a possible role of the sex chromosomes in disease progression. Three suggestive autosomal susceptibility QTL were also mapped in the chicken genome, with several orthologues of human systemic sclerosis candidate genes located within two of these loci.

## Genetisk analys av autoimmuna och metabola egenskaper hos kyckling

De första husdjuren tämjdes för ungefär 9,000-12,000 år sedan och vi har sedan dess valt ut och avlat på de djur som vi tycker har fördelaktiga egenskaper. Detta kallas selektiv avel (Clutton-Brock, 1995). Genom denna avel har våra husdjur fått intressanta och användbara egenskaper. Dessutom gör det dem till intressanta modeller för att studera hur gener reglerar dessa egenskaper. Selektiv avel används även i renodlat forskningssyfte, då man ofta väljer ut och avlar på de djur som har ovanliga fenotyper som påminner om, t ex mänskliga sjukdomar, och därmed skapas vad som kallas en genetisk djurmodell.

Studier av hur förändringar i arvsmassan påverkar organismer samt hur egenskaper ärvs från en generation till nästa kan ge oss en ökad förståelse kring olika egenskapers genetiska reglering, som t ex sjukdomar och produktionsegenskaper hos våra husdjur. Ökade kunskaper kring dessa egenskaper kan både hjälpa oss att utveckla nya mediciner för både djur och människor, samt att öka produktiviteten hos våra husdjur. Många sjukdomar och produktionsegenskaper regleras av en eller flera gener samt miljöfaktorer, vilket vanligtvis gör att det är svårare att finna de gener som reglerar egenskapen. En utmaning för forskare är därför att kartlägga de gener som reglerar variationen hos viktiga komplexa egenskaper inom medicin och husdjursaveln.

Forskare använder statistiska genetiska modeller för att bestämma geners additiva och dominanta effekter samt eventuella interaktionseffekter (epistasi) mellan loci. Additiva effekter är den del av genernas verkan som kan förklaras med skillnaden mellan de olika homozygoterna, medan dominanta effekter utgörs av heterozygotens avvikelse från medelvärdet av respektive homozygoter. En epistatisk effekt innebär att en gens effekt påverkas av alleler i ett annat lokus.

Det övergripande målet med denna avhandling är att använda och testa nya statistiska metoder på experimentell data, för att kartlägga gener som reglerar metabola och autoimmuna egenskaper hos kyckling.

## Bakgrund

De flesta metabola och autoimmuna egenskaper har visat sig vara komplexa, eftersom studier har visat att många gener, såväl som miljöfaktorer, påverkar egenskaperna.

En ökad förståelse kring metabola egenskapers genetiska reglering är viktig inom husdjursförädlingen, då viktiga metaboliska egenskaper som tillväxt och fettsammansättning har stor ekonomisk betydelse inom livsmedelsproduktion. En ökad förståelse för dessa egenskaper kan identifiera nya gener av vikt för framtida avelsprogram.

Forskning har visat att många autoimmuna sjukdomar är kopplade till genetiska effekter, infektioner och/eller miljöfaktorer. Autoimmunitet innebär att kroppens immunförsvar, som vanligtvis skyddar kroppen från främmande substanser, blir överreaktivt och attackerar kroppens egna celler. Ungefär 3% av jordens befolkning lider av någon autoimmuna sjukdom. I denna avhandling har jag studerat systemisk scleros (SSc), som är en systemisk sjukdom, vilket innebär att den attackerar flera organ, vävnader och celler, autoimmun sjukdom. Systemisk scleros ger ärrbildning på huden samt på interna organ som till exempel matstrupe, lungor, njurar och hjärta. Ett tydligt yttre kännetecken för sjukdomen är ärrbildning på fingerspetsarna.



För att kartlägga de gener som reglerar de komplexa egenskaper vi studerat har vi använt QTL-kartering. QTL står för "quantitative trait locus" och är en benämning för en kromosomregion som innehåller en eller flera gener som påverkar en komplex egenskap. Principen för QTL-kartering är att man först mäter egenskapen av intresse hos varje individ i den populationen man vill studera. Samma individer genotypas också för genetiska markörer (variabla positioner i DNA) spridda över genomet. Sedan använder man sig av statistiska test för att finna eventuella associationer mellan variationen av egenskapen och de genetiska markörernas genotyper.

## Sammanfattning av studierna

### Metabola egenskaper (artikel I-III)

För att studera genetiken bakom metabola egenskaper användes en korsning mellan två selektionslinjer där kycklingar har selekterats för hög, respektive låg kroppsvikt vid 56 dagars ålder under drygt 50 år (artikel I-III). Selektion har lett till att de två linjerna idag (generation 54) uppvisar mycket stora viktskillnader. I snitt väger högviktlinjen tolv gånger mer än lågviktlinjen. Tidigare studier av Carlborg *et al* (2006), har visat att ett nätverk bestående av fyra samverkande QTL förklarar 45% av viktskillnaden mellan dessa selektionslinjer.

I artikel I undersökte vi om, och hur, nätverket reglerar andra metabola egenskaper som t ex bröstmuskelvikt, buk fett och benvikt. Först inkluderade vi fler genetiska markörer i QTL-regionerna och bekräftade nätverkets effekt på kroppsvikt vid 56 dagars ålder. Vi visade att nätverket ökar kroppsvikten genom en symmetrisk viktökning hos övriga uppmätta organ. Ett av QTL paren visade sig även påverka den relativa bröstmuskelvikten samt buk fettsvikten.

I artikel II medverkade jag till att generera, samt analysera, en nio generationers "advanced intercross line" (AIL). För att generera en AIL så korsas individer från F<sub>2</sub>-generationen med varandra för att generera en F<sub>3</sub> population, o s v, tills man

uppnår önskat antal korsningsgenerationer. Målet var att finkartera de QTL-regioner som studerats tidigare. QTL-regionerna från originalstudien (Carlborg *et al.*, 2006) är stora och eftersom många gener ligger i regionerna är det svårt att veta vilken eller vilka av dessa gener som påverkar kroppsvikten. Genom finkarteringen kunde vi minska regionernas vidd, vilket skapar bättre förutsättningar för att hitta de gener som påverkar egenskaperna.

När vi bekräftat och finkarterat nätverket i AIL populationen designade vi ett avelsprogram för att generera en selekterad återkorsning ("Introgression line"). Denna typ av återkorsning syftar till att föra över specifika gener eller QTL från en donatorlinje till en mottagarlinje. Först korsas de två linjerna för att generera en F<sub>1</sub>-population. F<sub>1</sub>-populationen återkorsas sedan till mottagarlinjen, varpå selekterade avkommor återigen återkorsas till mottagarlinjen. Denna procedur upprepas ett önskat antal generationer. Målet med studien var att generera individer med låglinjealleler i QTL-regionerna och höglinjealleler i resten av genomet. En F<sub>1</sub>-generation mellan hög- och låglinjen återkorsades därför tillbaka till höglinjen och avkommor med högst frekvens låglinjealleler i QTL-regionerna valdes ut med hjälp av genetisk markörinformation, för att sedan återkorsas till nästa generation. Vi avlade fram tre generationer av återkorsningar. Individer från den tredje generationen korsades med varandra för att generera även individer som är homozygota för både hög- och låglinjealleler i de studerade QTL regionerna

#### Autoimmuna egenskaper (artikel IV)

För att identifiera gener som reglerar systemisk scleros använde vi en kycklinglinje selekterad för systemisk sclerosliknande symtom (UCD-200, artikel IV). Denna selektionslinje återkorsades med röd djungelhöna och avkommorna analyserades med tre olika typer av QTL-karteringsmodeller; en generaliserad linjär regressionsmodell, en varianskomponentmodell samt en modell för parvis epistasi.

Studien visade en signifikant könsskillnad i sjukdomsfrekvens; tuppar insjuknade med högre frekvens än honor. Vi fann tre QTL-regioner i höns-genomet med association till sjukdomen. I två av dessa QTL-regioner ligger flera gener som redan tidigare har visat sig vara kandidatgener för systemisk scleros hos människor. Detta resultat indikerar att UCD-200 hönsen är en användbar modell för att få fram ny kunskap om den genetiska regleringen bakom systemisk scleros.



## References

- Abplanalp, H., Gershwin, M.E., Johnston, E. & Reid, J. (1990). Genetic control of avian systemic sclerosis. *Immunol Genet.* 31, 291-295.
- Andersson, L. & Georges, M. (2004). Domestic-animal genomics: Deciphering the genetics of complex traits. *Nat Rev Genet* 5(3), 202-212.
- Besnier, F. & Carlborg, Ö. (2009). A genetic algorithm based method for stringent haplotyping of family data. *BMC genet* 10:57.
- Besnier, F., Wahlberg, P., Rönnegård, L., Ek, W., Andersson, L., Siegel, P.B., *et al.* (2011). Fine mapping and replication of QTL in outbred chicken advanced intercross lines. *Genet Sel Evol* 43(1):3.
- Beyer, C., Schett, G., Distler, O. & Distler, J. (2010). Animal models of systemic sclerosis: prospects and limitations. *Arthritis Rheum* 62(10), 2831-2844.
- Bouchard, G., Paigen, B. & Carey, M.C. (2000). In *Symposium in honour of Gustav Paumgartner* (eds Gerbes, A. *et al.*) 97-101.
- Broman, K., Wu, H., Sen, S. & Churchill, G. (2003). R/qt1: QTL mapping in experimental crosses. *Bioinform* 19, 889-890.
- Burt, D. W. (2005). Chicken genome: Current status and future opportunities. *Genome Res* 15(12):1692-1698.
- Carlborg, Ö. (2002). New methods for mapping Quantitative Trait Loci. Doctorat thesis. Acta Universitatis Agriculturae Suecia, Veterinara 121, Swedish University of Agricultural Sciences, Uppsala, Sweden.
- Carlborg, Ö. & Haley, C. (2004). Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* 5, 618-625.

- Carlborg, Ö., Jacobsson, L., Ahgren, P., Siegel, P. & Andersson, L. (2006). Epistasis and the release of genetic variation during long-term selection. *Nat Genet* 38, 418–420.
- Churchill, G.A. & Doerge, R.W. (1994). Empirical threshold values for quantitative trait mapping, *Genet* 138, 963-971.
- Clutton-Brock, J. (1995). A natural history of domestic animals. *Cambridge: Cambridge Univ Press. Second edition.*
- Crooks, L., Nettelblad, C. & Carlborg, Ö. (2011). An improved method for estimating chromosomal line origin in QTL analysis of outbred lines. *G3: Genes, Genomes, Genetics* 1:57-64
- Cutolo, M., Suli, A., Seriola, B., Accardo, S. & Masi, A.T. (1995). Estrogens, the immune response and autoimmunity. *Clin Exp Rheumatol* 13(2), 217-226.
- Darvasi, A. & Soller, M. (1995). Advanced intercross lines, and experimental population for fine genetic mapping. *Genet* 141(3), 1199-1207.
- Dietrich, H.M., Cole, R.K. & Wick, G. (1999). The natural history of the obese strain of chicken – an animal model for spontaneous autoimmune thyroiditis. *Poult Sci* 78, 1359-1371.
- Dunnington, E.A. & Siegel, P.B. (1996). Long-term divergent selection for eight-week body weight in White Plymouth Rock chickens. *Poult Sci* 75, 1168-1179.
- Eriksson, J., Larson, G., Gunnarsson, U., Bed-hom, B., Tixier-Boichard, M., Strömstedt, L., *et al.* (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet* 4(2): e1000010. doi:10.1371.
- Falconer, D.S. & Mackay, T.F.C. (1996). Introduction to quantitative genetics, fourth edn. *Harlow: Addison Wesley Longman.*
- Fernando, R.J. & Grossman, M. (1989). Marker-assisted selection using best linear unbiased prediction. *Genet Sel Evol* 21, 467-477.
- Finjeman, R.J., De Vries, S.S., Jansen, R.C. & Dermat, P. (1996). Complex interactions of new quantitative trait loci, Sluc1, Sluc2, Sluc3 and Sluc4 that influence the susceptibility to lung cancer in the mouse. *Nat Gen* 14, 465-467.
- Fridolfsson, A.K., Cheng, H., Copeland, N.G., Jenkins, N.A., Liu, H.C., Raudsepp, T., *et al.* (1998). Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc Natl Acad Sci USA* 95(14), 8147-8152.
- Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S. & Kondo, N. (1994). One subspecies of red jungle fowl (*Gallus gallus*

- gallus) suffices as the matriarchic ancestor of all domestic breeds. *Proc Natl Acad Sci USA* 91(26), 12505-12509.
- Gama, L.T., Smith, C. & Gibson, J.P. (1992). Transgene effects, introgression strategies and testing schemes in pigs. *Anim Prod* 54, 427-440.
- Gershwin, M.E., Abplanalp, H., Castles, J., Ikeda, R., van der Water, J., Eklund, J., *et al.* (1981). Characterization of a spontaneous disease of white leghorn chickens resembling progressive systemic sclerosis (scleroderma). *J Exp Med* 153(6), 1640-1659.
- Goldgar, D.E. (1990). Multipoint analysis of human quantitative genetic variation. *Am J Hum Gen* 47, 957-967.
- Green, P., Falls, K. & Crooks, S. (1990). Documentation for CRI-MAP, version 2.4. *Washington School of Medicine, St Louis*.
- Groen, A.F & Smith, C. (1995). A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *J Anim Bred Genet* 112, 161-170.
- Hafler, D.A. & Weiner, H.L. (1989). MS: a CNS and systemic autoimmune disease. *Immun Today* 10(3), 104-107.
- Haley, C.S. & Knott, S.A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69, 315-324.
- Haley, C., Knott, S. & Elsen, J-M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136(3), 1197-1207.
- Hall, R. & Stanbury J.B. (1967). Familial studies of autoimmune thyroiditis. *Clin Exp Immunol* 2, 719-725.
- Hedges, S. B. (2002). The origin and evolution of model organisms. *Nat Rev Genet* 3(11):838-849.
- International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695-716.
- Jacobsson, J., Park, H-B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, P.B. *et al.* (2005). Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. *Genet Res* 86, 115-125.
- Jensen, J. (1989). Estimation of recombination parameters between a quantitative trait locus (QTL) at two marker gene loci. *Theor Appl Genet* 78(5), 613-618.

- Kahalefp, M.B. & Leroy, E.C. (1999). Autoimmune and vascular involvement in systemic sclerosis (SSc). *Autoimmunity* 31(3), 195-214.
- Ke, X., Durrant, C., Morris, A.P., Hunt, S., Bentely, D.R., Deloukas, P. & Cardon, L.R. (2004). Efficiency and consistency of haplotype tagging of dense SNP maps in multiple samples. *Hum Mol Genet* 13(21), 2557-2565.
- Kuhn, A., Haust, M., Rauland, V., Weber, R., Verde, P., Felder, G. *et al.* (2010). Effect of bosentan on skin fibrosis in patients with systemic sclerosis: a prospective, open-label, non-comparative trial. *Rheum* 49(7), 1336-1345.
- Li, Z., Pinson, S.R., Park, W.D., Paterson, H.A. & Stansel, J.W. (1997). Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* 145, 453-465.
- Liu, G., Dunnington, E.A. & Siegel P.B. (1994). Responses to long-term divergent selection for eight-week body weight in chickens. *Poult Sci* 73, 1642-1650.
- Long, A.D., Mullaney, S.L., Mackay, T.F. & Langely, C.H. (1996). Genetic interaction between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in *Drosophila melanogaster*. *Genetics* 144, 1497-1510.
- Lynch, M. & Walsh, B. (1998). Genetics and analysis of quantitative traits. *Sunderland: Sinauer Associates, Inc.*
- McCanlies, E., O'leary, L.A., Foley, T.P., Kramer, M.K., Burke, J.P., Libman, A. *et al.* (1998). Hasimoto's thyroiditis and insulin dependent diabetes mellitus: differences among individuals with and without thyroid function. *J Clin Endocrinol Metab* 83(5), 1548-1551.
- Mendel, G. (1865). Experiment in Plant Hybridization (English version). In Bateson W. 1909. Mendels principle of Heredity. London: *Cambridge University Press.*
- Nelson, RM., Shen, X. & Carlborg, Ö. (2011). qtl.outbred : interfacing outbred line cross data with the R / qtl mapping software. *BMC res notes* 4, 154.
- Nguyen, V., Sgonc, R., Dietrich, H. & Wick, G. (2000). Endothelial injury in internal organs of University of California at Davis line 200 (UCD 200) chickens, an animal model for systemic sclerosis (Scleroderma). *J Autoimmun* 14(2), 143-149.
- Park, H., Jacobsson, L., Wahlberg, P., Siegel, P. & Andersson, L. (2006). QTL analysis of body composition and metabolic



- traits in an intercross between chicken lines divergently selected for growth. *Physiol Genomics* 25, 216-233.
- Podolin, P.L., Wilusz, M.B., Cubbon, R.M., Pajvani, U., Lord, C.J., Todd, J.A., *et al.* (2000). Differential glycosylation of interleukin 2, the molecular basis for the nod ldd3 type 1 diabetes gene? *Cytokine* 12 (5), 477-482.
- Prokunina, L. (2004). Strategies for identification of susceptibility genes in complex autoimmune diseases. Doctoral thesis. Acta Universitatis Upsalensis, 1333, The faculty of medicine, Uppsala University, Sweden.
- Rönnegård, L., Besnier, F. & Carlborg, Ö. (2008). An improved method for quantitative trait loci detection and identification of within-line segregation in F2 intercross designs. *Genetics* 178(4), 2315-2326.
- Schmid, M., Nanda, I., Hoehn, H., Scharl, M., Haaf, T., Buerstedde, JM., *et al.* (2005). Second report on chicken genes and chromosomes 2005. *Cytogenet Genome Res* 109(4), 415-479.
- Smith, D. & Germolec, D. (1999). Introduction to immunology and autoimmunity. *Environ Health Perspect* 107, 661-665.
- Smyth, J.R.Jr. (1989). The Smyth line chicken: a model for autoimmune amelanosis. *Poult Biol* 2, 1-19.
- Siegel, P. & Dunnington, E.A. (1987). Selection for growth in chickens. *Crit Rev Poult Biol* 1, 1-24.
- Simmons, D. (2008). The use of animal models in studying genetic diseases: transgenesis and induced mutation. *Nat Education* 1(1).
- Symula, D.J., Frazer, K.A., Ueda, Y., Deneffe, P., Stevens, M.E, Wang, Z-E. *et al.* (1999). Functional screening of asthma QTL in YAC transgenic mice. *Nat Genet* 23, 241-244.
- van de Water, J. & Gershwin, E.M. (1985). Animal model of human disease. Avian Scleroderma, an inherited fibrotic disease of white leghorn chickens resembling progressive systemic sclerosis. *Am J pathols* 120(3), 478-482.
- Von Tschermak, E. (1951). The rediscovery of Gregor Mendel's work. *J Hered* 42(4), 163-171.
- Wick, G., Andersson, L., Hala, K., Gershwin, ME., Selmi, C., Erf, GF., *et al.* (2006) Avian models with spontaneous autoimmune diseases. *Adv immunol* 92, 71-117.

- Wolf, E., Lahm, H., Wu, M., Wanke, R. & Hoeflich, (2000). Effects of IGFBP-2 overexpression *in vitro* and *in vivo*. *Pediatr Nephrol* 14, 572–578.
- Wolf, J.B., Brodie III, E.D. & Wade, M.J. (2000). Epistasis and the evolutionary process. *Oxford University Press*.
- Öklu, R., Hesketh, TR., Metcaife, JC. & Kemp, PR. (1998). Expression of alternatively spliced human latent transforming growth factor binding protein-1. *FEBS Lett* 435, 143–148.

## Acknowledgements

The studies in this thesis were carried out at the Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala. We gratefully acknowledge funding from, EURYI, VR, FORMAS and swedish foundation for strategic research. The SLU fund for internationalisation of postgraduate studies provided travel grants for my visit at QIMR, Australia and one of my travels to Purdue University, USA. KSLA funded my travels to IOWA state university, Virginia Tech State University and Purdue University.

I would like to express my sincere gratitude to all of you who have helped and supported me during my time as a PhD student. There are some of you that I like to give special credit:

First of all I would like to thank my supervisors:

*Örjan Carlborg*, thank you for believing in me, for your encouragement and most of all for your great support. Thank you for all the sincere care you show for everyone in the group and for creating a social, friendly, relaxed, but still challenging working environment.

*Susanne Kerje*, for the collaboration with the autoimmune projects and for always showing me your appreciation and encouragement.

*Lucy Crooks*, thank you for all the hours of staring at non-working programs, for all your excellent contributions in matters of evaluation and analysis of data, for making me

survive the cold in Poland and especially for all the help and support during the last days of finishing up.

The whole (present and former) *Computational genetics* group for making everyday at work fun, interesting and challenging! For sharing your great knowledge of life and science, for all great cakes, food and wine and for putting up with me on Fridays. It has been a great privilege to work within a group of so qualified and supportive researchers!

*Francois*, for all the programming exercises but mostly for all the fun, ping-pong, French wine, babyboxes and for an amazing barbecue in France.

*Anna*, for all your great ideas and for always being ready to help.

*Ronnie*, for all your beautiful graphs, long discussion about food and religion, and for letting me be crazy during working hours. Looking forward to visit South Africa with you and Carina...one day...

*Mats "mr google"*, for always having candy in store and for all the expertise on programming, politics and language. Also for all the help with my last paper.

*Marcin*, for sharing the love of Phi and not setting my house on fire.

*Xia*, for your amazing brain and for letting me use it sometimes.

*Xidan*, for always being friendly and kind.

*Stefan*, the best room mate, for our collaboration with the introgression, for running advice, life questions and just a lot of laughter.

*Lars*, for always answering all my stupid (BUT FUN) statistics question and for the times in beautiful Djura

*Beyong Woo Kim*, for introducing me to R and for all the nice Korean gifts.

*Linda Roest*, my first master student, for all the fun and for a work well done! Also thank you for the nice visit in Alphen ad Reijn... or wherever we were...

*Andreas*, for not making me freak out totally...

*Paul Siegel*, For letting me stay at "Siegel Hilton" during my trip to Blacksburg and for all your help and feedback during these 4 years. Also, thank you for showing me mountain lake and how a real hamburger should taste.

*Uppsala University: Leif, Olle, Olov, Anna-Stina and Johan* for the collaboration with the Autoimmune projects.

*Per W and Lina J*, for all the help in the beginning of my project.

*Linneaus centre for bioinformatics- Olle and Jacke*, for making LCB a cool place!

*Purdue University: William Muir and Ashok* for the collaboration with the introgression line as well as nice dinners in Purdue. *Fred and Barry* för all the help at the poultry farm. Also, big thanks to my students at Purdue, especially to *Dustin* who worked many many hours with me and the smelly chickens without any complaining.

*IOWA University: Jack Dekkers and Sue Lamont*, for the invitation to IOWA and for giving me a nice time in Ames.

*INRA, France: Elisabeth Blesboi and Isabelle Grassaue*, for showing me avian cryopreservation and always responding to my questions.

*Queenslands institute of Medical Reserarch: Peter Vissher* and his group for sharing your great knowledge and for letting me be part of the group for 3 month. Thanks to *Joseph* for inviting me to Australia and for all the nice trips around the world. Thanks to *Enda*, for all the help with Plink and Merlin.

All the people at the *Department of Animal breeding and Genetics, SLU*, for making everyday at work interesting and fun.

*Lina C, Sandra, Anne and Helena-* för att vi har ju nästan roligt jämt!

*Marie E* for being a nice friend and for all your encouragement.

*Per*, för att du låtit mig leka med underbara Ebba.

*Jessika, Moncan, Elisenda och Anna O* för hård kroppstortyr varje tisdag

*Jennie*, för allt stöd under mina sista stressade dagar, tack!

*Erik Bongcam* for fun and interesting discussions about work, life and everything inbetween.

*DJ*, thank you for all your suport and sincere interest in my projects.

*Helena Pettersson*, världens bästa sekreterare. Tack för att du alltid har funnits till hands med ett leende, hur många reseräkningar jag än har kommit in med.

*My family*: Tack till *Pappa Alf* för att du alltid har trott på mig och låtit mig lära genom mina egna misstag. Tack för att du ställt upp i vått och torrt, *Mamma Bodil* för allt du gav mig och för att du lärt mig att allt är möjligt, önskar så att du var här. *Linda* för att du är världens bästa syster och ett speciellt tack för hjälpen med cryoförsöken och 2 roliga veckor i USA. *Eva*, *Ingemar*, *Pontus* och *Patrik* för att ni alltid funnits där för mig. Speciellt tack till *Pontus* för korrekturläsning av mina artiklar samt för roliga veckor i USA!

*My new extra family in Uppsala: Erika, Maria, Pia och Anders*. Tack för att ni gett mig ett riktigt hem i Uppsala, tillgång till underbara hästar och för att jag fått vara en del av eran underbar familj. Tack *Pia* för alla gånger du har lyssnat och frågat, det har betytt mer än du kan tro. *Erika* och *Maria*, det finns inte plats att skriva hur viktiga ni är för mig. Tack för allt ni är!

*My Australian family: Gavin, Sibel, Kristel, Michelle, Tonsi and Joseph!* Thank you all for inviting Alfie to your homes, for making my time in Australia amazing and fun, for all the fun trips with climbing, swimming, walking and surfing. And not to forget all the GTs in the big sofa! Extra tack till *Erika* och *Anna Aminoff* för en fab semester i Oz och för tiden ni spenderade med mig i Brisbane!

*Barbro*, för allt stöd under en svår tid med många beslut, varav det viktigaste tog mig till Uppsala.

Thanks to all my other supporting, amazing and fun friends, you all know who you are =)