THE GENETIC ARCHITECTURE OF THE CHICKEN GENOME

by

GEORGINA ABENA ANKRA-BADU

(Under the Direction of SAMUEL E. AGGREY)

ABSTRACT

Traits of economic importance are controlled by several genes and the environment. The detection of quantitative trait loci (QTL) and the genes underlying them are therefore important for the improvement of these traits. A study was conducted to locate putative candidate genes on GGAZ by orthologous comparison of QTL regions on GGAZ with the mouse and human genomes. Primer sequences from markers flanking QTL regions were blasted against the chicken genome using BLASTN (http://www.ensembl.org). Forty six chicken genes together with 91 mouse and 60 human genes were identified in this study. The annexin A1 gene, follistatin and neuronal acetylcholine receptor gene (nAChR) were some of the genes identified in this work. The nAChR gene is located at a QTL region for abdominal fat and could be used a therapeutic agent for feed intake and obesity. A second study was conducted to detect QTL regions for growth and skeletal traits in an F_2 population selected for high and low growth. QTL for age-related body weight (BW), shank length (SL) and shank diameter (SD) were localized in 695 individuals. A pleitropic QTL on GGA4 explained 7-11% of the variance and affected BW at 5 to 9 weeks, and SL and SD at 9 weeks. A male –specific BW QTL was detected on GGA3 at 173 cM. The QTL on GGA4 had the strongest on SL and SD, and explained 18% and 21%, respectively of the variance. A third study located QTL for carcass composition and fat on GGA

2, 3, 4,5,10 and 26. Differentially expressed genes in the QTL regions included SOD and fat-1. The Spot 14 gene is associated with abdominal fat in chickens. The fourth study was conducted to characterize a major QTL region on GGA4 and to identify candidate genes in this region by CpG island detection and comparative mapping. One hundred and nine known genes and 179 CpG islands were located at this locus. Six putative novel genes were identified by blasting genes from 23 orthologous species against the chicken genome. A putative ortholog of the rhotekin gene was detected. Rhotekin is a housekeeping gene with acetylcholine activity.

INDEX WORDS: CpG islands, candidate genes, QTL

THE GENETIC ARCHITECTURE OF THE CHICKEN GENOME

by

GEORGINA ABENA ANKRA-BADU

BSc, University of Ghana, 1999

MSc, The University of Georgia, 2003

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2006

THE GENETIC ARCHITECTURE OF THE CHICKEN GENOME

by

GEORGINA ABENA ANKRA-BADU

Major Professor:	Samuel E. Aggrey
Committee:	Gene Pesti
	Paul Schliekelman
	Romdhane Rekaya
	Larry McDougald

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2006

DEDICATION

To the Lord who has brought me this far

ACKNOWLEDGEMENTS

I would like to thank the good Lord for giving me the grace and strength to complete my PhD program. I would also like to thank my major professor, Dr. Aggrey, for his advice, encouragement and support. Thank you Dr. Aggrey, for giving me the chance to work with you. I am also grateful to the members of my advisory committee, Dr. Gene Pesti, Dr. Paul Schliekelman, Dr. Romdhane Rekaya and Dr. Larry McDougald for agreeing to be on my committee in spite of their busy schedules. I would like to thank the Faculty, staff and graduate students of the Poultry Science department for providing a great working environment. I have had a good five years because of all of you. To Aarti who started out as my office mate and ended up being a dear friend, thank you for being you. I am grateful to Dr. Jessie Kissinger, Dr. Russell Malmberg, Dr. Mary Bedell and Dr. Alex Feltus for their patience and their willingness to share their knowledge in genetics and bioinformatics with me. Finally, I am so grateful to God for my boyfriend Justice, my loving family and my wonderful friends for their constant love and support. God bless you all.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Prior to the advent of biotechnology in animal production, the breeding value of an individual was determined by its phenotype or the average of its relative's performance. However, these criteria may not be good estimators of an individual's breeding value because they are influenced by chromosomal rearrangements and environmental effects. Furthermore, traits of agricultural importance are complex traits that are influenced by several genes and their interaction with the environment. The detection of the gene loci affecting these traits and the subsequent use of marker assisted selection (MAS) in animal breeding may overcome these problems and ensure that the best combination of genetic material is transmitted to the offspring. Nonetheless, the identification of these gene loci, also known as quantitative trait loci (QTL) is only the first step in the detection of genes associated with economic traits. It is also important to identify the causative genes which may be utilized in MAS for the improvement of agricultural traits. Therefore, the ultimate goal of QTL analysis is to identify potential candidate genes that underlie QTL regions. Several methods of gene identification which are complementary to each other have been developed for this purpose. Some of these techniques include fine mapping, comparative genomics, candidate gene analysis and the identification of CpG islands. Fine mapping is a crucial intermediary step in the process of candidate gene analysis because it narrows down the QTL interval and reduces the number of potential genes to a workable number. Comparative mapping is based on the

fact that most species still show large regions of conserved synteny after diverging from a common ancestor millions of years ago. It is a useful tool when genes of known function correspond to loci controlling genes of interest (Pflieger, 2001). Candidate gene analysis involves the use of statistical tests to evaluate the association between variants of the causative gene (s) and polymorphisms in the trait of interest. These methods may however be limited to studies where prior knowledge about the position and function of the gene is known. Therefore, CpG island analysis, which does not require prior knowledge of a gene, can be used as an additional tool for the detection of quantitative trait genes. CpG islands are clusters of CpG nucleotides with a high GC content which are primarily associated with the 5' ends of genes. They have been used for the detection of genes in several species due to their distinctive nature.

The identification of QTL and the subsequent identification of the causative genes have several significant implications. First, MAS which involves the use of polymorphic markers associated with QTL or those directly located in the gene can be used for the incorporation of desirable genes into animal breeding. Secondly, the detection of the causative genes could promote the use of transgenic technology for the improvement of quantitative traits (Falconer and Mackay, 1996; Borevitz and Chory, 2004). Finally, the identification of candidate genes can lead to the simultaneous improvement of traits that have a negative correlation with fitness and lead to an overall improvement in agricultural production.

Research Objectives:

1. To identify QTL regions for growth, body weight and body composition on the chicken genome
2. To detect candidate genes in QTL regions on GGAZ by comparative mapping
3. To characterize the genomic landscape of a major QTL region on GGA4

CHAPTER 2

REVIEW

The identification of the genes underlying complex traits is challenging because these characters are controlled by several genes and gene loci which are also influenced by the environment. A major approach to overcoming this problem involves the detection and mapping of the underlying gene loci, usually referred to as quantitative trait loci (QTL). QTL analysis involves associations between quantitative traits and markers (Doerge et al., 1997) and requires the use of well designed experiments and statistical methods ranging from simple ANOVA to the more complex Bayesian methods. However, the detection of QTL is only the first step in gene identification because they usually span very large regions of the chromosome which must be narrowed down in order to identify the actual causative trait genes. Fine mapping, which localizes these regions more precisely, serves as an important step towards achieving this goal. Several methods of fine mapping which depend on the population of interest have been developed. These techniques may reduce QTL regions to about 1 cM in order to clone causative genes and pave the way for gene identification methods like candidate gene analysis, comparative mapping and CpG island detection. The gene identification stage must then be followed by validation to ensure that polymorphisms of these genes are associated with variations in the trait.

1. Quantitative trait loci analysis: A Historical perspective

1.1 Early studies on QTL

Studies by Fisher and other researchers on the inheritance of quantitative traits paved the way for the basic idea of QTL in the early part of the 20th century (Weller, 1997). The first QTL study was by Sax (1923) who detected trait loci in the common bean *(Phaseolus vulgaris)*. However, major research into QTL inheritance did not begin until the 1980s due to the lack of markers that were polymorphic enough to allow for optimal QTL detection. The discovery of DNA markers in the 1980s led to an increase in QTL studies on species of scientific and economic interest. From 1960 to 1980 several new methods for QTL analysis were developed though these were mostly related to experimental populations with fixed allele effects in inbred lines (Weller, 1997). Neimann-Soressen and Robertson (1961) initiated studies on outbred populations when they developed the half-sib method of QTL analysis in cattle. These researchers who were the first to use blood groups rather than morphological markers also developed the chi-squared and ANOVA methods of statistical analysis in QTL mapping. Jayakar (1970) and Haseman and Elston (1972) later developed the likelihood function for estimating QTL parameters and recombination.

1.2 QTL studies in chickens

One of the first theories about genetic markers and their association with economic traits in chickens was proposed by Serebrovsky and Petrov (1930). These workers developed the signal gene concept which indicates that a signal gene which does not influence a trait by itself may serve as a marker if it is in the vicinity of a gene that

influences that trait. Additionally, the closer the distance between the signal gene and the economic trait gene, the more accurate the prediction of this influence. Major studies on chicken QTL may have began in the 1990s with the detection of QTL for fat and body weight (Plotsky et al., 1990; Dunnington et al., 1990). Subsequent QTL studies have also focused on growth, body weight and egg traits as well as other traits of agricultural importance.

1.3 QTL studies on growth and fat traits

Van Kaam et al. (1999) conducted a QTL study on growth and feed efficiency using a full sib interval mapping approach. They located four significant QTL for these traits on chromosomes 1, 2 and 4 for body weight and feed intake. The most significant QTL, which was located on chromosome 1, was associated with both growth and feed intake. Sewalem et al. (2002) detected a significant QTL for body weight at 3, 6 and 9 weeks on chromosome 13. A second QTL affected body weight at two ages on chromosomes 1, 2, 4, 7 and 8. The genetic effects of these QTLs were mostly additive and accounted for between 0.2 and 1.0 phenotypic standard deviations. Zhu et al. (2003) generated an F2 intercross from a cross between two commercial broiler lines with various degrees of resistance to coccidiosis. Five parameters including oocyst shedding were evaluated in this study and a QTL for oocyst shedding which had mainly additive effects was associated with chromosome 1 (LOD=3.46). Three possible growth QTL were also detected on chromosomes 1, 6 and 8. A study on an F2 population obtained from crossing broiler and layer lines detected QTLs for fat traits (Ikeobi et al., 2002). Chromosomes 3, 7, 15 and 28 were associated with QTLs for abdominal fat weight while chromosomes 3, 7 and 13 contained skin and subcutaneous fat QTLs. Generally, the percentage of the variance explained by the QTL ranged from 3-5%. Jennen et al. (2004) identified

significant QTL body weight for percent abdominal fat on chicken chromosome 1. QTL for abdominal fat weight at different ages were detected on chromosome 4.

1.4 QTL studies on egg traits

Tuiskula-Haavisto et al. (2002) located 14 genome-wide significant and six suggestive QTL on chromosomes 2, 3, 4, 5, 8 and Z. Significant QTL for egg white thinning and eggshell strength were identified on chromosomes 2 and Z respectively. Several highly significant QTL for body weight, egg weight and feed intake were located in the same area on chromosome 4. The most significant QTL, which was for body weight explained about 26% of the variance. Chromosome Z was associated with several egg-related traits. Several QTL for body weight and egg characteristics have been identified from a cross between White Leghorn and Rhode Island Red chickens (Sasaki et al., 2004). Significant QTL were located for body weight on chromosomes 4 and 27. QTL for egg-related traits were also located on chromosomes 4 and 11 and a trait locus for age at first egg was detected on chromosome Z. Schreiweis et al. (2005) detected several QTL for egg color, egg and albumen weight, percent shell, body weight and egg production. Eleven QTL exceeded the 5% genome-wise significance level and 64 tests from 17 linkage groups exceeded the 5% chromosome-wide significance threshold. A QTL region on chromosome 2 influenced egg shell color at 35 and 55 weeks. Chromosome 4 had several significant QTL for egg and albumen weights, percentage of shell, egg production and body weight. The QTL for egg and albumen weights were the most significant, explaining 11-19% of the total variance. These studies indicate that chromosome 4 contains several QTL associated with traits of economic importance.

Other studies on QTL have located chromosomal regions associated with tonic immobility, restraint and fearful behavior (Kerje et al., 2003), social tendency (Schutz et

al., 2002), coccidiosis (Zhu et al., 2003), Marek's disease (Vallejo et al., 1998; Lipkin et al., 2002), resistance to salmonella (Kaiser et al., 2002), ascites (Rabie et al., 2004) and meat quality (van Kaam et al., 1999).

2. Quantitative trait loci analysis: Experimental design and statistical methods

2.1 Experimental design and mapping population

The design for a QTL experiment is dependent on the species of interest and the experimental population. These designs include the backcross (BC) and F_2 designs in inbred lines, and the half-sib and full sib families in natural populations (Lynch and Walsh, 1998; Liu, 1998). Backcross and F_2 populations are usually used for crosses that are genetically uniform and are fixed for allele effects. In the backcross population two parents with opposite allele effects are mated to produce the F_1 population. The F_1 offspring are then mated to either parent to produce the backcross progeny. These populations are however not efficient when dominance effects need to be estimated therefore the more efficient and powerful F_2 population which is developed by intercrossing the F_1 offspring may be used as an alternative. On the other hand, individuals in the F_2 population that are heterozygous at a given locus provide no information for the estimation of additive effects thus other populations like recombinant inbred lines and double haploid lines which create a highly homozygous background may be used for increased power (Lynch and Walsh, 1998; De Vienne and Cause, 2003). In spite of the above observations the BC design may be more effective than the F_2 cross when there is dominance at both the marker locus and QTL, depending on which parent is used in the backcross. In populations where inbred lines cannot be established for

practical or ethical reasons, oubred populations with segregating QTL are used for QTL mapping. The detection of QTL effects in this type of population is more complex. (Lynch and Walsh, 1998). First, linkage information can only be obtained when the parent is heterozygous at both the marker and the QTL thus homozygous parents are non-informative. Also while inbred lines have only two segregating alleles in the population there are usually more than two alleles segregating in the outbred population. Thirdly, marker-trait associations in outbred populations are not fixed. In inbred lines marker-QTL phase is always known thus generating the required amount of linkage disequilibrium. Conversely, in a segregating population the M marker could be associated with either the Q or q allele in outbred populations. Consequently, it is necessary to establish the QTL phase in this population by establishing marker trait associations for each parent. In spite of these limitations QTL mapping in outbred populations is useful because it provides information about the segregation of QTL within a population. Experimental populations in outbred lines consist of groups of relatives like the half sib and full sib families. Large half sib families are usually used in animal breeding while small full sib families are more practical in human genetics due to biological limitations (Lynch and Walsh, 1998).

A half sib and F_2 design were used for the detection of QTL for backfat and susceptibility to Marek's disease in pigs and chickens respectively (de Koning et al., 1999; Vallejo et al., 1998). Trait loci for fat were detected in a chicken population using the line-cross, half-sib and backcross designs (Abasht et al., 2006).

2.2 Methods of QTL detection

i) Single Marker Analysis: The simplest method of detecting a QTL involves the use of a single marker. Animals are divided into groups based on their genotype and traits measurements are taken for each group. Statistical analyses like a simple ANOVA is then conducted to determine whether there are differences between the means for the each group. A difference between the means is an indicator that a QTL is present. A major disadvantage of this method is its inability to distinguish between a QTL location and its effect thus one is unable distinguish between a QTL of small effect and one that is actually a QTL of large effect located at a great distance from the marker (De Vienne and Causse, 2003). Interval mapping which involves the use of two markers has been developed to separate these two confounding effects.

ii) Interval mapping: This method is the most widely used in QTL analysis and is based on the assumption that there is only one QTL between two markers with a recombination rate r. The genotypic value of each genotype is a function of the effect of the QTL and its location and the rates of recombination between the QTL and its flanking markers. The benefits of interval marker methods over single marker analysis are minimal for intervals smaller than 20cM but increases as the interval becomes larger (De Vienne and Causse, 2003). Though interval mapping is very efficient in distinguishing between the QTL position and effect, it is not very efficient in the detection of more than one QTL. Therefore two QTL in coupling phase are detected as one phantom QTL (De Vienne and Cause, 2003). Multiple marker methods like composite interval mapping (CIM) and multiple interval mapping (MIM) have been develop to correct these problems.

iii) Composite interval mapping: This method was developed separately by Jansen (1993) and Zeng (1993). It combines interval mapping with multiple linear regression. In this method one tests for the putative QTL by using others markers as covariates to control for other QTL and to reduce the residual variance in order to improve the possibility of detecting the QTL (Kao et al., 1999). Since this method still makes use of one QTL the multiple interval mapping method, which utilizes a multiple QTL model, was developed to improve the detection of multiple QTL.

iv) Multiple interval mapping: This method was developed by Kao et al. (1999) and is based on the likelihood method of estimating genetic parameters. It involves the use of all the linked markers on the chromosome at the same time resulting in a single analysis (Lynch and Walsh, 1998). Multiple interval mapping makes allowances for missing data and is also efficient for mapping multiple QTL and the interactions between them (Kao et al., 1999; Broman, 2001).

2.3 Statistical methods in QTL mapping

The simplest method of QTL analysis involves the use of linear models. These include the simple t-test, ANOVA and regression methods (Doerge, 2001; Liu, 1998). Linear models utilize marker means for the analysis and require the use of data with a normal distribution. The simple t-test is used to detect the difference between two marker means e.g. backcross populations while ANOVA methods are used to detect differences between more than two marker means. With these two methods individuals are separated into different classes based on their marker genotype which will be M_1M_1 and M_1M_2 for example in the backcross population.

The expected difference between the two means is given as:

$$\mu_{M1\,/\,M1} - \mu_{M1\,/\,M2} = (1\text{-}2r_{MQ})\,\sigma$$

where σ is either the difference between the means for the two genotypes or the sum of the additive and dominance effect (a + d) and r is the recombination fraction between one marker and the QTL. Hence a trait mean difference of zero either implies that there is no genetic effect or that there is no linkage between the marker and the QTL (Liu, 1998; Doerge at al., 1997). The t test statistic for the test of difference between the means is given as:

$$t = \frac{\mu M1/M1 - \mu M1/M2}{\sqrt{S2(1/nm1/m1 + 1/nm1/m2}}$$

where $\mu_{M1\,/\,M1}$ and $\mu_{M1\,/\,M2,}$ are the sample means for the two marker classes, s^2 is the pooled sample variance and $n_{M1\,/\,M1}$ and $n_{M1\,/\,M2}$ are the sample means for the two marker classes. (Doerge et al., 1997)

The linear regression method regresses the trait value on the marker genotypes which are represented by dummy variables and can be applied to both single marker and interval mapping methods. The model is given by:

$$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j$$

where y_j is the trait value for the jth individual in the population and x_j is the dummy variable that takes on a value of 1 or 0 depending on whether the marker has the genotype M_1 / M_1 or M_1 / M_2. β_0 is the intercept and β_1 is the slope or the expected difference between the trait values for the marker classes. Thus a null hypothesis $H_{0:}\ \beta_1 = 0$ indicates that there is no linkage between the marker and the QTL and r = 0.5 (Liu, 1998; Doerge,

2001). Linear models are simple methods for QTL detection but they only make use of marker means and may not be robust when assumptions of normality are violated. Other methods of analysis like the maximum likelihood method have been developed to include all the information from the marker-trait distribution (Lynch and Walsh, 1998).

The procedure for estimating the maximum likelihood estimates is as follows: The probability of the observed data is estimated given the QTL parameters μ_A, μ_B and σ^2 and the recombination fraction r. This probability, Pr (data / μ_A, μ_B, σ^2, r), is referred to as the likelihood. The maximum likelihood estimates are the probability values for which the parameters achieve their maximum (Broman, 2001).

The test statistic for this analysis is the log likelihood ratio which is given as:

$$G = -2\text{Ln}\left[\frac{L(\mu 1, \mu 2, \sigma 2, r = 0.5)}{L(\mu 1, \mu 2, \sigma 2, r)}\right]$$

where $\mu 1$, $\mu 2$ are the estimates of the trait means and σ^2 is an estimate of the variance. The numerator represents the log likelihood of the null hypothesis of no difference between the means while the denominator is the alternate hypothesis that a QTL is present (Doerge, 2001). The conventional test statistic is the logarithm of odds score (LOD) which measures the probability of for the presence of a QTL location z.

The score is given as follows:

$$\text{LOD} = \log 10 \left\{ \frac{\Pr(\text{data} / \text{QTL at z}, \mu 1, \mu 2, \sigma)}{\Pr(\text{data} / \text{no QTL})} \right\}$$

This measures the strength of the evidence that there is QTL at a location z against the evidence of no QTL. A high LOD implies strong evidence for a QTL and the standard cutoff LOD score is 3 (Broman, 2001). Large LOD scores correspond to low p-values which indicate that the null hypothesis must be rejected.

The QTL techniques that have been described fall under the frequentist method which becomes unmanageable as the number of QTL increases. Also, these methods may result in an overestimation of QTL effects because they depend on the number of markers (Hoeschele and Vandraden, 1993). Though several solutions have been proposed, the use of Bayesian methods by Markov Chain Monte Carlo (MCMC) methods may provide the best answer to problems associated with multiple QTL. This is due to the fact that unlike the frequentist approach which uses hypothesis testing, the Bayesian approach expresses all the results in terms of the posterior distribution of the parameters in the model given the data. The number of QTL and the QTL positions are defined by considering the marginals of the posterior distribution (Maliepaard et al., 2001). This kind of model has been used successfully for mapping QTL in outbred populations (Hoeschele and Vanraden, 1993; Uimari and Hoeschele, 1997).

2.4 Statistical issues in QTL detection

i) Power of QTL experiments: The power of a QTL research depends on the type I and II errors. The type I error (α) is the probability of observing a factor when there is none while the type II error (β) is the probability of not observing a factor when it is exists. In QTL studies the type II error refers to the probability that an experiment will fail to detect a QTL when it is present. This implies that the power of a QTL test (1- β) is the probability that a QTL will be detected by an experiment (de Vienne and Causse, 2003).

Thus although it is important to reduce the possibility of false positives and negatives, the values of α and β should not be selected at the expense of the power of the experiment. On the other hand, lower values of α may be selected in experiments with a large number of analyses as the standard levels of 0.05 and 0.01 will result in a large number of false positives. For a given sample size the probability of detecting a QTL when it is present also depends on the additive effect and the within class variance. Therefore, a small difference between alleles or a large distance between the QTL and the marker may lead to a reduction in the power of the test.

ii) Population size and marker density: A large population size and marker density increases the power of QTL detection thus large sample sizes are required for the detection of QTL of small effect while smaller sizes may be suffice for QTL with large allelic differences. Additionally, for a given sample size it is better to increase the number of genotypes rather than the number of replications per genotype in order to increase the power of detection.

iii) LOD score: This score, which is used in interval mapping, is applied to traits with continuous variation and a normal distribution. Therefore, traits such as disease symptoms which have skewed distributions may produce abnormal values and may be more suitable for non-parametric tests.

3. Quantitative trait loci analysis: Detection of candidate genes

The ultimate goal of QTL analysis is to detect the underlying genes associated with the trait of interest. The discovery of the causal genes at these loci is however difficult because QTL regions are very large and may contain thousands of putative candidate

genes. Hence, several methods of gene identification have been developed to narrow down these regions, locate the genes responsible for the QTL by determining their relationship with polymorphism of the trait and provide a framework for marker assisted selection. These techniques, which complement each other, may be used in combination with or as a follow up to fine mapping. Identification of the actual genes affecting quantitative traits is necessary for the application of technology to traits of agricultural importance. Some of the methods that have been utilized in QTG detection are fine mapping, comparative mapping, positional cloning, candidate gene analysis, microarray analysis and the detection of CpG islands.

3.1 Fine Mapping

Fine mapping / high resolution mapping of QTL regions can be broadly categorized into linkage methods and association based methods depending on the population of interest.

i) Linkage analysis

Fine mapping techniques based on linkage analysis can be categorized into genome wide based methods and locus based methods (Darvasi, 1998). Genome wide based strategies include the use of backcross, F2 and half sib populations as well as intercross lines and heterogenous stock. Locus based strategies, which include selective genotyping, recombinant progeny testing, interval specific congenic strains and advanced intercross lines are required to narrow down the analysis to a single QTL (Darvasi, 1995; Darvasi, 1998).

1) Selective genotyping

This technique requires the establishment of a backcross or an F_2 population. However, the only individuals that are phenotyped are those that are recombinant at the QTL interval. This is based on the theory that only the recombinants at the QTL region contribute to mapping accuracy. The selective genotyping is done in stages and at each stage the total number of animals phenotyped is reduced by 1/2r (1-r) for an F_2 population and 1/r for a BC population, where r is the proportion of recombination between the markers marking the QTL interval. This method only requires 2 generations for fine mapping. However, very large samples are necessary as the resolution increases thus it is most appropriate when the resolution required does not go below 5cM.

2) Recombinant progeny testing

Progeny carrying an obvious recombinant chromosome at the QTL region being studied are crossed to one of the individuals with the parental genotype to determine the location of the gene compared to the recombination point. A reduction in the confidence interval from one value (x) to another (y) requires the use of y/x individuals each with a recombination at one of the y/x intervals covering the y interval. The expected number of F_2 animals that will be screened to detect these y/x individuals is given by

$$N_s = \frac{50}{x} \sum_{i=1}^{\frac{y}{x}} 1/i$$

where Ns is the total number of F_2 required. This method requires only three generations for fine mapping and is effective for the estimation of dominant effects. However, it requires a large number of samples.

3) Interval-specific congenic strains (ICSC)

This system is similar to recombinant progeny testing because a number of F_2 individuals are tested to detect y/x recombinant individuals with recombinations equally distributed within the y-cM interval. The major difference is that, the ICSC progeny is mated several times to the parental strain to eliminate alleles from all other QTL affecting the trait. The progeny are then intercrossed and homozygotes for the recombinant are selected to establish the strains. The different stages of selection are conducted with DNA markers thus the number of generations required is reduced. This technique has also the advantage of requiring few progeny even for QTLs of small effect. Also, the pool of ISCS progeny can be used for additional phenotyping. It is however not efficient for studies with dominance effects.

4) Recombinant inbred segregation test (RIST)

This design employs the high resolution in recombinant strains and utilizes it in QTL mapping. The QTL containing interval is reduced from y-cM to x-cM by selecting y/x recombinant strains with recombination equally distributed in the y-cM interval. Since a single recombinant strain is capable of having more than one recombination in the region of interest fewer progeny are required in this design. The setup for this design is as follows: The F_2 or BC is phenotyped and genotyped with a few markers. The progeny is then crossed with the parents (P_1 and P2) to produce two types of populations $F_{1,1}$ and $F_{1,2}$ in the case of the F_2 population or two different sets of backcross populations in the BC. These are then intercrossed or backcrossed to produce the RIST-F_2 and RIST-BC progeny respectively. Two categories of F_2 are produced, the $F_{2,1}$ and $F_{2,2}$. Similarly, there are two types of RIST-BC, BC_1 and BC_2. The $F_{2,1}$ or BC_1 progeny are genotyped

with markers in the region where the P_2 alleles are present in the recombinant strain. On the other hand, the $F_{2,2}$ or BC_2 progeny is genotyped with markers in the region where the P_1 alleles are present in the recombinant strain. Since the QTL has already been mapped to this location it will segregate in one population but not the other. The analysis of the two populations will determine the population where the QTL is segregating and will locate the QTL above or below the recombination point. The QTL will then be located to the required interval by overlapping the results of all the recombinant strains. The RIST-BC population is more efficient for dominance effect while RIST-F_2 is more effective for determining additive effects. This method requires only two generations and does not require a very large sample size even for QTL of small effect. However, it requires the availability of recombinant strains with recombination in the areas of interest.

5) Advanced intercross lines

This technique utilizes the statistical power derived from the use of inbred lines and saturated genetic maps and involves the continuous intercrossing of a population to reduce linkage disequilibrium. The advanced intercross population is developed by the sequential and random intercrossing of inbred lines for several generations. As a result of this, recombinant events required for fine mapping are obtained in a relatively small population over several generations. Advanced intercross lines can be used for the reduction of a QTL confidence interval without an increase in the number of genotyped individuals. It can also be used for high resolution mapping of QTL with small effects. However, the sample size for this design must not be less than a hundred.

ii) Association studies

Association studies, where candidate genes have been directly tested for their association with traits have also been used in fine mapping. These studies which utilize family based and population based designs are ideal for humans and other natural populations whose breeding cannot be controlled by the experimenter. They increase resolution of the QTL interval and identify at-risk alleles (Bull et al., 2005). One major disadvantage of this method is that a mutation at a locus of interest may be associated with more than one haplotype. Moreover, it requires controls to minimize confounding due to population stratification and other environmental factors (McPeek, 2000).

Fine mapping via various methods have been used for the reduction of QTL intervals and the identification of the causative genes. Candidate genes for muscular dystrophy have been identified on GGA2 by linkage analysis (Yoshizawa al., 2004). Wang et al. (2003) identified genes associated with lung tumors in mice using advanced intercrossed lines. Selective genotyping has been used to localize a genetic marker associated with body weight in chickens (Feng, 1998) and genes associated with disease resistance have been detected with congenic and recombinant strains (Bacon et al., 2000). Gunnarson et al. (2006) narrowed down a large QTL for growth in chickens by association mapping. Selective backcrossing was done for White Leghorn females and males that were heterozygous for the growth QTL from the F3 and F4 generations and progeny testing was carried out to test for sires that were heterozygous for the QTL. Association studies were then conducted between the phenotype and genotype at close marker positions in the F4 and F5 generations.

3.2 Comparative mapping

Comparative mapping identifies potential novel genes by detecting homologous chromosomal segments in distantly related species (Johannson et al., 1995). For example, well developed maps like those of the human and mouse can be compared to the relatively poorly map of the chicken genome. Homologous segments of these species are then identified and a quantitative trait gene may be localized based on data from the well mapped species. Comparative genomics also allows for the identification of evolutionary trends among species because it is based on the fact that most species show large regions of conserved synteny though they diverged millions of years ago. Comparative mapping has been used successfully in the detection of genes associated with complex traits in several species. Yu et al. (1997) identified several genes in a gene-poor region of HSA 21 associated with Down's syndrome by the construction of microclones and the subsequent comparison of these clones with several databases. Nanda et al. (2000) detected the chicken double sex and mab related transcription factor 1 (DMRT1) by locating its homolog in humans. Thirty seven genes were located by comparative mapping of the chicken genome with HSA 1, 4 and 9 (Suchyta et al., 2001). These workers observed that there were high levels of conserved synteny between the genomes in spite the rearrangements in the order of the genes. Orthologous comparison of QTL regions of GGAZ with HSA and MMU also identified a total of 197 putative candidate genes (Ankra-Badu and Aggrey, 2005). About 78% of these genes were detected in the mouse and human genomes. Some potential candidate genes like the nicotinic acetylcholine receptor and the annexin 1 gene may be associated with abdominal fat and egg related traits in chickens. Stoll et al. (2006) located putative regions for human hypertension by

comparative mapping. Comparison between 67 QTL regions for hypertension in rats and humans predicted 26 genomic regions that may contain genes associated with human hypertension. These workers indicated that similar loci have also been identified in the mouse and may provide information for additional functional studies. Comparative mapping may however be subject to the definition of homology thus in studies where the QTL interval is narrowed down to about to an appreciable level, positional cloning methods may be used.

3.3 Positional cloning

This technique is utilized in studies where the physiological, molecular basis and the function of the gene are not known (de Vienne and Causse, 2003). It involves high resolution mapping to narrow down the QTL region to about 0.3cM (Falconer and Mackay, 1996). Refining the QTL position to this interval can be achieved by the fine mapping methods that have been described in the previous section. After high resolution mapping, candidate genes in the critical region can be cloned and tested to determine whether there is a causal relationship between the mutation of the gene and the trait of interest. Identification of the positional candidate gene involves literature and similarity searches, and the evaluation of expression patterns of the gene in the parental strains. Transgenic or knockouts techniques (Korstanje and Paigen, 2002) can also help to identify the candidate gene and determine its function. This method is technically laborious (Falconer and Mackay, 1996), is only effective for QTL of large effect and cannot be used to resolve a QTL to a single locus. Furthermore, it requires a large sample size and can only be utilized in a species with a small genome (de Vienne and Causse, 2003). Its advantage is that it narrows down the interval in the search for functional genes

and it can be utilized even when the function of the gene is unknown (Stratil and Geldermann, 2004). Positional cloning has been utilized in the identification of the genes involved in chromosomal aberrations in certain human papillary renal cell carcinomas (Weterman et al., 1996).

3.4 Candidate gene approach

This approach involves the identification of a gene of known function which is subsequently analyzed to determine whether its polymorphism is associated with variations in the trait of interest (de Vienne and Causse, 2003). The main advantage of this method is that unlike positional cloning it does not require the approximate region of a QTL of interest (Falconer and Mackay, 1996). However, association between variant alleles and a trait does not always imply causation (Stratil and Geldermann, 2004). Moreover, detailed knowledge of the biological characteristics of the traits and biochemical properties and developmental pathways of the potential candidate genes are required for successful candidate genes analysis (Falconer and Mackay, 1996). In spite of these shortcomings, this approach has been used effectively for the identification of allelic variants which are associated with polymorphisms in a trait. An example is the relationship between polymorphism in the NRAMP1 gene and susceptibility to tuberculosis in humans (Bellamy et al., 1998). There have also been reports of the association between the alleles of Apolipoprotein E and total serum cholesterol and heart disease in humans (Davignon et al., 1988; Nelson et al., 1999). Candidate gene analysis has been utilized successfully in chickens. Bennett et al. (2006) determined the relationship between the Vitamin D receptor (VDR), osteopontin (SPP1), insulin-like growth factor 1 (IGF1) and insulin (INS) genes; and bone, egg and growth traits in

chickens. They detected significant associations between variations in VDR and bone mineral content. They also indicated that polymorphisms in the IGF1 and INS genes were associated with body weight at 5 weeks and 55 weeks respectively. A study to identify associations between the growth hormone (GHR), gonadotrophin-releasing hormone receptor (GNRHR) and neuropeptide Y (NPY) and reproductive traits detected a dominance effect for NPY on age at first egg and an additive effect of GNRHR on the number of double yolks (Dunn et al., 2004). In spite of the usefulness of this technique the process of selecting a gene for candidate gene analysis may be further simplified and accelerated by microarray analysis (Korstanje and Paigen, 2002).

3.5 Microarray analysis

Microarray studies involve the detection of differences in gene expression and have been combined with QTL analysis to reduce the number of putative candidate genes to a more manageable number (Wayne and McIntyre, 2002). This technique is very useful where *a priori* candidate genes based on a biological model are not available (Wayne and McIntyre, 2002). Aitman et al. (1999) identified the Cd36 gene which underlies a QTL for defects in glucose and fat metabolism by using a combination of microarray studies, congenic mapping and radiation hybrid mapping. They reported that mice which over expressed this gene had reduced blood lipids. Moreover, its deficiency is associated with insulin resistance and may be important in the pathogenesis of human insulin-resistance. Karp et al. (2000) also used a combination of QTL analysis, microarray technology and single-nucleotide polymorphism for the detection of a gene encoding for complement factor 5 (C5), a susceptibility gene for allergen-induced airway hyperresponsiveness in mice models of asthma. They indicated that a blockade of the C5a receptor prevents the

production of interleukin 12 which reverses asthma. Wayne and McIntyre (2002) reduced the number of genes located in a QTL region for ovariole number by combining QTL analysis with microarray technology. Initial studies on this trait by QTL analysis identified 5,286 genes which were reduced to 548 by fine mapping. Microarray technology detected genes that were differentially expressed thus reducing the number of putative causative genes for ovariole number to thirty four. Tabakoff et al. (2003) used similar techniques together with selective breeding to locate a protein that is associated with tolerance to the ataxic effects of alcohol. In a study on drought tolerance in rice, Nguyen et al. (2004) identified 14 QTL regions associated with root traits on chromosomes 1, 2, 4, 5, 6, 7, 8, 9, 10 of the rice genome. Microarray technology was used to identify candidate genes at these loci which are drought-related and may be implicated in drought resistance. Liu et al. (2001) used a combination of microarray techniques and genetic on a sample of 15 genes. They reported that one of the differentially expressed genes is associated with resistance to Marek's disease. One problem with the microarray approach is the failure to detect genes that are either tissue-specific or are weakly expressed (www.informatix.org/silver). This implies that some genes may go undetected therefore other methods of identification which do not depend on gene expression must also be utilized.

3.6 Identification of CpG islands

CpG islands are unmethylated regions of the genome with a high concentration of CpG nucleotides and GC content. Due to their distinctive structure and their associated with the 5' ends of genes, CpG islands are useful for the detection of genes that cannot be detected by other methods. Due to their association with genes, CpG islands have been

utilized in the estimation of the number of genes on the mouse and human genomes (Antequera and Bird, 1993). CpG islands have also been used for the categorization of genes based on their tissue specificity (Larsen et al., 1992; Ashikawa, 2002). Other studies on CpG islands have involved the characterization of QTL regions and the identification of underlying genes. Weber et al. (1991) conducted a study to determine the precise position of a region associated with Huntington's disease (HD). They detected 15 CpG islands and indicated that those sequences associated with CpG islands detect cross-species conservation and provide a framework for the identification of genes associated with HD. Maestrini et al. (1992) used 19 CpG islands as probes for the construction of a physical map of genes on the X chromosome. They reported that CpG islands are clustered in a 2-Mb region and provide information relevant to the identification of candidate genes which are associated with diseases that have been mapped in this region. Lee et al. (2006) located 8 genes and 33 CpG islands in a QTL regions associated with growth in the pig. They observed that the regions with a high concentration of CpG also had a high number of genes. CpG islands have been used for the isolation of full length cDNAs due to their association with the transcriptional ends of genes (Cross et al., 1999). Additionally, CpG islands isolated from the MBD1 gene contained the first exon and regulatory sequences of the gene. Gillespie et al. (1991) identified candidate genes for autosomal dominant polycystic kidney disease by locating CpG islands in a region associated with the disease. The CpG islands were then used as markers to identify genes associated with this disease.

3.7 Testing and Verification of candidate genes

An important stage in candidate gene identification involves validation which verifies the association between the putative genes and the trait of interest. This process involves testing for the association between gene polymorphisms and phenotypic differences at the population level, and transformation with transgenic constructs (Thornsberry et al., 2001; Pflieger, 2001; De Vienne and Causse, 2003).

i) Association studies / Map cosegregation: Once putative candidate genes are identified they are screened by testing for the co-segregation or statistical correlation of their alleles with the QTL (Pflieger, 2001). In cosegregation tests, genetic map positions of putative genes and trait loci can be compared. For association studies, candidate genes are sequenced to test for polymorphism in their coding and regulatory regions. Statistical analysis must be conducted to test for the correlation between the alleles of the genes and the trait in question. A variant of a gene of interest is said to be associated with a trait if it occurs at a significantly higher frequency among affected individuals compared with the control (Lander and Schork, 1994). A statistical correlation or a map co-segregation may not necessarily imply a causal relationship thus where possible transgenic techniques must be applied.

ii) Genetic transformation: This involves the introduction of a sense construct of the candidate gene into a phenotype that is deficient for the trait of interest. A restoration of the functional phenotype confirms the association between the gene and the trait. Another method that is commonly used in plants involves the introduction of an antisense construct into a non-deficient phenotype. This approach is used when the expression of the candidate gene leads to an inhibition in the expression of a functional gene. The

association between the candidate gene and the trait is verified if the transformants develop a deficiency in the trait.

Other methods like the physiological analysis of the expression of the gene have been proposed but these are less conclusive than transgenic techniques (Pflieger, 2001).

CHAPTER 3

IDENTIFICATION OF CANDIDATE GENES AT QUANTITATIVE TRAIT LOCI ON CHICKEN CHROMOSOME Z USING ORTHOLOGOUS COMPARISON OF CHICKEN, MOUSE AND HUMAN GENOMES[1]

---

[1] Georgina A. Ankra-Badu and S. E. Aggrey Submitted to In Silico Biology, 2005

ABSTRACT

This study was undertaken to identify novel candidate genes at quantitative trait loci (QTL) on chicken chromosome Z (GGAZ) by comparing orthologous regions of chicken, human and mouse genomes. Primer sequences from marker flanking QTL positions (https://acedb.asg.wur.nl/) were obtained from http://www.iastate.edu/chickmap and blasted against the chicken genome (http://www.ensembl.org) using BLASTN. The best matches were those with the highest score, lowest E-values and highest percent identity. Orthologous regions in mice and humans, together with genes located on or around those loci were identified using the Ensembl website. Forty-six chicken genes, 91 mouse genes and 60 human genes associated with QTL on GGAZ were identified in the current study. Among the most promising candidate genes for egg production and egg shell quality are annexin A1 (*ANXA1*), osteoclast stimulating factor (*OSF*), thrombospondin-4 (*THBS4*), programmed cell death proteins (*PDCD*), follistatin (*FST*), growth hormone receptor (*GHR*), interferon (*IFN*) α and ß. The chicken IFN α and ß were located on GGAZ around position 13,000,000 bp on the draft chicken sequence map. The neuronal nicotinic acetylcholine receptor (*nAChR*) is located at a QTL region for abdominal fat (GGAZ 25483091 bp). Nicotine is an agonist at the *nAChRs* and has been shown to decrease lipolysis and triglyceride uptake, thereby reducing net storage in adipose tissue. Therefore, the *nAchRs* could be used as therapeutic targets for regulating feed intake and obesity. This study has identified 197 putative candidate genes in probable QTL regions of chicken chromosome Z.

INTRODUCTION

Traditional methods for genetic improvement in farm animals and poultry usually depend on Mendelian principles and quantitative genetics. With these methods the breeding value of an individual is determined either by its phenotype or by the average performance of its relatives (Falconer, 1960; Falconer and Mackay, 1996). The expected genetic gain would depend on the additive genetic variance associated with the trait. When the heritability of a trait is high, the phenotype becomes a good predictor of an animal's breeding value and artificial selection results in rapid gains. However, genetic gains in lowly heritable traits have been limited. Breeding values on sires for sex-limited traits such as egg or milk production can only be determined from data from female progeny, and that requirement can prolong the generation interval.

Knowledge about genes that affect traits allows the breeder to manipulate the genome of an animal to ensure that the best combinations of alleles in the parental population are transmitted to the progeny. Marker assisted selection, which involves the use of molecular markers for genetic improvement, is one method to achieve this. A genetic marker is a phenotypically recognizable trait that can be used to identify a genetic locus, linkage group or recombination event (Aggrey and Okimoto, 2003). Molecular markers can also be used to generate genetic maps of many species and combined with trait measurements to determine locations on chromosomes harboring major genes or quantitative trait loci (QTL) that affect traits of economic importance (Van Kaam *et al*., 1998). Trait loci generated by genome scans do not provide information on the linkage phase between the marker alleles and QTL, the number of genes within the QTL location nor the nature of gene interactions at that location. Therefore, selection on QTL or

markers associated with QTL could result in less then expected gains. To gain insight into the molecular basis of QTL, it is desirable to identify the actual genes responsible for the QTL. Candidate gene analysis has been used as an alternative method in characterizing genes with a known function and their association with traits of economic importance (Kuhnlein *et al.,* 1997; Aggrey *et al*., 1999; Causse *et al*., 2004). Generally, the functional candidate gene approach is not adequate without the use of positional information from gene mapping studies (Haley, 1999). Examination of genes at QTL locations could lead to the identification of novel genes and could also narrow the number of genes to be analyzed.

Although the chicken genome is estimated to have about 20,000-23,000 genes (Hillier et al., 2004), only about 400 human gene orthologues have been mapped (Burt and Hocking, 2002). In spite of the tremendous achievement of a first draft, there are still some gaps, and these are particularly important to the Z chromosome. Comparative mapping of the chicken with other vertebrates has the potential to unravel information about new potential genes and to identify homologous chromosomal segments in distantly related species (Johansson *et al*., 1995). Through comparative mapping, potential candidate genes for hypertension in rats were uncovered for the same disease in humans (Rapp *et al*., 1989; Jacob *et al.,* 1991). O'Brien and Nash (1982) mapped 31 cat genes whose homologs have been previously mapped in humans and mice.

The chicken genome has been found to have high levels of synteny with the human and mouse genomes even though these species diverged from each other more than 300 million years ago (Burt and Hocking, 2002). Information derived from the human genome can be used to resolve unanswered questions involving the chicken

genome (Suchyta *et al.,* 2001). The chicken *double sex and mab related transcription factor 1* (*DMRT1*) for example, was isolated by identifying its homolog in humans (Nanda *et al*. 2000). Ladjali-Mohammed *et al.* (2001) localized four homeobox genes in a study that identified new sections of conservation between humans and chicken. Smith *et al*. (2000) mapped the chicken *rip associated ICH-1 homolog protein with a death domain* (*RAIDD*) by identifying its homolog in the mouse. Therefore, additional chicken genes can be identified through comparative mapping with other species.

The objective of this study was to identify novel candidate genes at QTL locations on *Gallus gallus* chromosome (chr) Z (GGAZ) by comparing orthologous regions of chicken, human and mouse genomes. Genetic markers in such novel genes could be used to aid in the genetic improvement of sex-limited traits or traits with low heritability.

MATERIALS AND METHODS

Information on QTL on GGAZ was obtained from https://acedb.asg.wur.nl/. Data on loci and markers on GGAZ were selected by generating a list of all the markers on the chromosome using information from http://www.genome.iastate.edu/chickmap. The consensus 2000 chicken linkage map (http://www.genome.iastate.edu/chickmap), which is a combination of mapping data from the East Lansing, Compton and Wageningen chicken populations, was used to locate the position of each marker. Comparative mapping was done by blasting the primer sequence of the markers against the chicken genome sequences in the Sanger Institute website (http://www.ensembl.org) using the basic local alignment search tool (BLASTN) for comparing a nucleotide query sequence against a nucleotide sequence database. A score is given for matching and mismatching

nucleotides and gaps. The total score is given by obtaining the sum of all matches, mismatches and gap penalties for sequence. The E-value or expect score is the number of different values that are equivalent to or better than the score that are expected to occur in a database by chance. The percent identities refer to the extent to which sequences are invariant (http://www.ncbi.nlm.nih/gov). A score of more than 45, percentage identity of greater than 70%, and an E-value less than 0.05 are considered to be significant (Pertsemlidis and Fondon, 2001; Jiang and Michal, 2003). The best matches were those regions that had the highest scores, lowest E-values and highest percentage identities. Each matching sequence was then compared with the mouse and human genome sequences to identify regions of homology. Information on genes at or around the QTL location on GGAZ and their respective homologs in humans and mouse were obtained from http://www.ensembl.org by identifying genes within homologous regions.

RESULTS AND DISCUSSION

Quantitative trait loci and their flanking markers on GGAZ are presented on Table 1. Loci for growth, body weight and abdominal fat were found at positions 22 (Kerje *et al.,* 2003), 96 (Sasaki *et al*., 2004), and 127cM (95% CI: 56-127 cM) (Ikeobi *et al*., 2002), respectively on the genetic map. QTL for age at first egg were located on positions 22 (Sasaki *et al.,* 2004) and 63-104 cM (90% C.I.: 65-137cM) (Tuiskula-Haavisto *et al.,* 2002). Quantitative trait loci for egg shell strength and thickness were also identified by Sasaki *et al.* (2004) at position 36 and 47 cM, respectively. Loci for other egg production traits (egg shell thickness, egg shell weight, egg number and egg weight) were identified

at positions 63-104 cM (Tuiskula-Haavisto *et al.,* 2002). Zhou *et al.* (2003) found a QTL for antibody response to *Brucella (B) abortus* on position 28.

Table 2 shows homologous regions between the QTL and the mouse and human genomes. The QTL regions on GGAZ show conserved synteny with mouse (MMU) chr 4, 13 and 19, and human (HSA) chr 5 and 9. The synteny between chicken and human for this QTL region is consistent with Nanda *et al.* (2002) who indicated that GGAZ and HSA 5 and 9 must have diverged from a common ancestor. Table 3 provides a list of the genes at probable chicken QTL regions and their homologs in mice and humans. Forty-six chicken genes, 91 mouse genes and 60 human genes were identified in this study and selected genes of interest were categorized by Gene Ontology (GO) annotation (Ashburner et al., 2000).

The QTL region for antibody resistance to *B. abortus* and egg shell strength contained *annexin A1* (*ANXA1*) and *nuclear orphan receptor ROR-β* (*RORB*). Annexins are involved in the biological processes of arachidonic acid secretion, cell cycle and signal transduction. They are calcium regulated phospholipid and membrane-binding proteins (Rescher and Gerke 2004). *ANXA1*, a member of the annexin family, is an important endogenous modulator of inflammation (Yona *et al.,* 2005) and is actively expressed in lymphoid tissues. However, its direct involvement in resistance against or susceptibility to *B. abortus* in chicken is not known. *ANXA1* could be a candidate gene for egg shell quality since it has calcium binding properties and is secreted in the epithelial and endothelial lining of the endometrium (Bedford *et al*., 2003). Other genes found in the QTL region for egg shell strength and other egg quality traits included the *osteoclast stimulating factor* (*OSF1*), *riboflavin kinase* (*RFK*) and the *guanine nucleotide*

binding protein (*GNB*). *OSF1* is a transcription factor (GO: 0003700). It enhances osteoclast formation and bone absorption through a cellular signal transduction cascade (Kurihara *et al*., 2001). In humans, Kurihara *et al.* (2001) suggested that the *OSF* interaction with *survival motor neuron* (*SMN*) could be important in a novel signaling cascade that induces stimulators of osteoclast formation. Dodds *et al*. (1995) reported that osteoclasts control the deposition of osteopontin, which is present in the egg shell, bone and other hard tissues (Gautron *et al*., 2001). Several chicken genes were found in the QTL region for egg traits, and they mostly control cell differentiation, embryonic development and immune response. They include *thrombospondin 4* (*THBS4*), *programmed cell death protein* (*PDCD*), *follistatin* (*FST*) precursor and *growth hormone receptor* (*GHR*) all of which play key roles in cellular and biological processes. The *THBS4* and *GHR* genes are also involved in molecular function and calcium binding and receptor activity, respectively. Orthologous regions in the mouse and human for the QTL harbor several *INF* α genes. *THBS*s are a family of related calcium binding glycoproteins found in the embryonic extracellular matrix (Tucker *et al*., 1995) that are associated with tissue genesis and remodeling. The *THBS4* promoter is similar to promoters of housekeeping, growth regulating, and other *THBS* genes which contain multiple GC box sequences and lacks a CAAT box. The presence of multiple E-box motifs is consistent with *THBS4* expression in muscle and cartilage, tendon and bone tissue (Newton *et al.,* 1999). *THBS4* could be a candidate gene for egg shell strength because it belongs to a group of calcium binding proteins that regulate tissue genesis (Lawler *et al.,* 1995). The human *PDCD* was found in the region homologous to egg related traits in chicken. PDCD is involved in the transcriptional regulation and biological processes of apoptosis.

It controls cell death in the female germline of several species and is believed to remove defective cells unable to develop after fertilization (Buszczak and Cooley, 2000). Programmed cell death ensures that viable eggs will receive nutrients, which could explain why this gene is associated with egg related traits. *FST,* a gene involved in female gonad development (GO: 0008585) and gametogeneis (GO: 0007276) suppresses the secretion of the *follicle stimulating hormone* (*FSH*) and reverses the effect of activin on oxytocin and progesterone thereby preventing degenerating effects on dominant follicles (Michel *et al*., 1993). *FST* has the ability to bind the pleiotropic growth and differentiation factor activin, thereby neutralizing activin action. This glycoprotein is potentially an important regulatory factor, capable of modulating autocrine and paracrine functions which would alter differentiation and development (Farnworth *et al*., 1995). *GHR* is associated with an array of production traits (Kuhnlein *et al.,* 1997) and *GHR* variants have been shown to be associated with reproduction, growth and immune response (Feng *et al.,* 1998). These workers reported that selection for egg production in white leghorns had led to a co-selection of a *GHR* variant.

With the exception of *IFN*-γ and -ω, the entire IFN family located on MMU4 is homologous to the QTL region for egg traits. *IFNs* are part of the immune system that controls resistance to viral and bacterial infections and are known to exist in chicken embryo (Isaacs and Lindenmann, 1957). Interferons are involved in extracellular space, defense response and cytokine activities. Kaspers *et al.* (1994) demonstrated that chicken IFN mediated the induction of MHC class II antigens on peripheral blood monocytes. In a study to determine the effect of interferons on chicks with Marek's disease, Volpini *et al*. (1996), demonstrated that the Marek's disease antigen is down regulated by IFNs and

other cytokines. Nanda *et al*. (1998) mapped chicken IFN1 and IFN2, which are homologous to *IFN-α* and *IFN-β* respectively, to the short arm of the GGAZ at position p2.2-p2.4. Since the chicken orthologs of mouse *IFN-α* and *IFN-β* are located between 13,001,112 and 13,001,639 bp on GGAZ, the chicken *IFN-α* and *IFN-β* could be located at this region with a high probability. The location of *IFN* in the QTL region for egg traits could imply that disease resistance traits are also associated with egg related traits. Since some genes of the IFN family are transcription factors that regulate antigen expression, mutations of these genes could affect other genes that control immune response in the chicken. Therefore polymorphisms in these genes could control both egg production and immunity.

A homolog of the zinc finger protein, *basonuclin 2* (*BNC2*), was found on MMU4 which is homologous to the QTL region for egg quality. *BNC2* is a transcription factor that maintains cell proliferation and prevents terminal differentiation (Vanhoutteghem and Djian, 2004). The strong conservation of *BNC2* among vertebrates strongly suggests an important function, presumably as a regulatory protein of transcription. *Claustrin* (*MAP1B*), a keratin sulphate proteoglycan and *collagen receptor* (*VLA-2 alpha chain*) could also be novel candidate genes for egg shell strength. The eggshell is an ordered structure comprised of calcium carbonate deposits onto an organic matrix. It is made up of a mineralized portion (95%) and the organic phase (3.5%) (Gautron *et al.,* 2001). The non-mineralized portion of the egg shell matrix is made up of collagen while the mineralized portion contains keratin sulphate proteoglycans (Dennis *et al.,* 2000). The proteoglycans and egg white proteins are involved in the nucleation of calcite crystals on the outer membrane of egg shell (Gautron, 2001). Chicken genes found at the QTL region

for abdominal fat weight were *sarcomeric creatine kinase (*CKMT2), *creatine kinase* (*CK*), *progesterone receptor binding protein* (*PGRBP*) and the *neuronal nicotine acetylcholine receptor* (*nAChR*). Mouse genes in the homologous regions included *fukutin* (*FKTN*), *olfactory receptors* (*OR*), and *tyrosine kinase* (*TXK*). *Fukutin* and *TXK* are cellular components with *TXK* having a molecular function as well. *CK* has molecular functions and is also involved in biological processes. Human genes in the homologous regions included *endosome associated protein* (*EA1*) and *CK*. *CK* is a muscle-specific enzyme that plays an important role in energy transfer by catalyzing conversion of creatine to creatine phosphate with the expenditure of ATP. *CK* activity is especially high in tissues with high-energy transfer (Sattler and Furll, 2004) and could affect abdominal fat deposition by increasing energy utilization. Steroid hormones affect adipose tissue metabolism (Pederson *et al.,* 2003). Pederson *et al.* (2003) showed that progesterone counteracts the action of glucocorticoids, which increase central accumulation of fat tissue. The *nAChR* is located at position 25483091 bp at the QTL region for abdominal fat. Neuronal *nAChR*s are distributed throughout the central and peripheral nervous system. In muscle, *AChR*s are found exclusively at the neuromuscular junctions and are responsible for mediating the effects of nicotine (Li *et al.,* 2003). Nicotine is metabolized extensively into a series of metabolites. The physiological effects of nicotine are produced through its agonist interaction with the acetylcholine receptor.

Li and Kane (2003) have shown that nicotine regulates appetite, body fat and weight gain in rats via upregulation of *hypocretin (orexin) neuropeptide precursor* (*HCRT*), *neuropeptide Y* (*NPY*), and *leptin* (*LEP*) in the forebrain areas. In the periphery, *LEP* is down-regulated while *uncoupling protein 1* (*UPC1*) is up-regulated after nicotine

administration. Nicotine has also been found to decrease lipolysis and triglyceride uptake hence reducing net storage in adipose tissue (Sztalryd *et al*., 1996). The effect of nicotine and its agonism at both neural and muscular *AChRs* has led to the assumption that *AchRs* could be therapeutic targets for regulating feed intake and obesity (Li *et al.,* 2003). Therefore the chicken it is possible that *AChRs* could be associated with obesity.

In avian species, the constitution of sex chromosomes is ZZ for males and ZW for females. The genes located on the Z chromosome in females follow a sex-linked inheritance, whereas their male counterparts follow Mendelian inheritance. Evaluation of genes on GGAZ should be performed separately for both sexes to account for sex-linkage because no allele is transmitted by the dam to female progeny. There are limitations in comparing the current genetic and sequence maps, because the resolution of the genetic map is modest. This places wide confidence intervals on marker locations and on recombination rates resulting in a non-linear concordance between location in bp and position in cM.

SUMMARY

Despite the limitations of the current genetic and genome sequence maps, the present study has demonstrated that comparative mapping can be utilized to identify novel candidate genes potentially associated with traits of economic importance. One hundred and fifty-one additional genes that could not be extracted from the chicken draft sequence were located through comparative mapping. Furthermore, differentially expressed genes from microarray experiments can potentially sort out the role of candidate genes that may be tightly linked together within the same QTL region.

However, before any gene is considered as a candidate for genetic improvement programs, fine mapping must first be performed to ensure that these genes are indeed located within the confidence interval of the QTL position, secondly genetic markers have to be developed and their association with traits must be demonstrated in a population segregating for variants of the candidate gene.

REFERENCES

Aggrey, S.E., Yao, J., Sabour, M.P., Lin, C.Y., Zadworny, D., Hayes, J.F. and Kuhnlein,U. (1999). Markers within the regulatory region of the growth hormone receptor gene and their association with milk related traits in Holsteins. J. Hered. **90**, 148-151.

Aggrey, S.E. and Okimoto, R. (2003). Genetic markers: Prospects and Application in Genetic Analysis. *In:* Poultry Genetics, Breeding and Biotechnology, Muir, W.H. and Aggrey, S.E. (eds), CABI Publishing, Oxon, U.K. pp. 419-438.

Ashburner, M., et al.; The Gene Ontology Consortium (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25-29.

Bedford, K.J., Richmond, I., Taylor, A.D. and Atkin, S.L. (2003). Lipocortin-1 expression within normal and abnormal endometrium. Endocrine Abstracts **5**,138.

Burt, D. and Hocking, P.M. (2002). Mapping quantitative trait loci and identification of genes that control fatness in poultry. Proc. Nutr. Soc. **61**,441-446.

Buszczak, M. and Cooley, L. (2000). Eggs to die for: cell death during Drosophila oogenesis. Cell Death Differ. **7**, 1071-74.

Causse, M., Duffe, P., Gomez, M.C., Buret, M., Damidaux, R., Zamir, D., Gur, A., Chevalier, C., Lemaire-Chamley, M. and Rothan, C. (2004). A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J.Exp.Bot. **55**, 1671-85.

Dennis, J.E., Carrino, D.A., Yamashita, K. and Caplan, A.I. (2000) Monoclonal antibodies to mineralized matrix molecules of the avian eggshell. Matrix Biol. **19**,683-692.

Dodds, R.A., Connor, J.R., James, I.E., Rykaczewsski, E.L., Appelbaum, E., Dul, E. and Gowen, M. (1995). Human osteoclasts, not osteoblasts, deposit osteopontin unto resorption surfaces: an in vitro and ex vivo study of remodeling bone. J. Bone Miner. Res. **11,** 1666-1680.

Falconer, D.S. (1960). Introduction to Quantitative Genetics. Oliver and Boyd, Edinburgh, UK pp 117-118.

Falconer, D.S. and Mackay, T.F.C. (1996). Introduction to Quantitative Genetics. Longman, Harlow, UK pp 356-376.

Farnworth, P.G., Thean, E., Robertson, D.M. and Schwartz, J. (1995). Ovine anterior pituitary production of follistatin in vitro. Endocrinology **136**, 4397-4406

Feng, X.P., Kuhnlein, U., Fairfull, R.W., Aggrey, S.E., Yao J. and Zadworny, D. (1998). A genetic marker in the growth hormone receptor gene associated with body weight in chickens. J. Hered. **89,** 355-359.

Gautron, J., Hincke, M.T., Mann, K., Panheleux, M., Bain, M., Mckee, M.D., Solomon, S.E. and Nys, Y. (2001). Ovoclyxin-32 a novel chicken eggshell matrix protein. Isolation, amino acid sequencing, cloning and immunocytochemical localization. J.Biol. Chem. **276,** 39243-39252.

Haley C. (1999) Advances in quantitative trait loci mapping. AgBiotechNet. http://agbio.cabweb.org.

Hillier et al., International chicken genome consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspective on vertebrate evolution. Nature **432,** 695-716.

Ikeobi, C.O.N., Woolliams, J.A., Morrice, D.R., Law, A., Windsor, D., Burt, D.W. and Hocking, P.M. (2002). Quantitative trait loci affecting fatness in the chicken. Anim. Genet. **33,** 428-435.

Isaacs, A. and Lindenmann J. (1957). Virus interference. I. The interferon. Proc.Roy.Soc. Lond. B.Biol. Sci.147, 258-267.

Jacob, H.J., Lindpainter, K., Lincoln, S.E., Kusumi, K., Bunter, R.K., Mao, Y., Ganten, D., Dzau, V.J. and Lander, E.S. (1991). Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell **67,** 213-224.

Jiang, Z. and Michal, J.J. (2003). Linking porcine microsatellite markers to known genome regions by identifying their human orthologs. Genome **46,** 798-808.

Johansson, M., Ellegren, H. and Andersson, L. (1995). Comparative mapping reveals extensive linkage conservation-but with gene order arrangements between the pig and the human genomes. Genomics **25,** 682-690.

Kaspers, B., Lillehoj, H.S., Jenkins, M.C. and Pharr, G. T. (1994). Chicken interferon mediated induction of major histocompatibility complex class II antigens on peripheral blood monocytes. Veterinary Immunol. Immunopathol. **44, 71**-84.

Kerje, S., Carlborg, O., Jacobsson, L., Schutz, K., Hartmann, C., Jensen, P. and Andersson, L. (2003). The two fold difference in adult size between the red jungle fowl and white leghorn chickens is largely explained by a limited number of QTLs. Animal Genetics **34,** 264-274.

Kuhnlein, U., Ni, L., Weigend, S., Gavora, J.S., Fairfull, W. and Zadworny, D. (1997). DNA polymorphisms in the chicken growth hormone gene: response to selection for disease resistance and association with egg production. Anim. Genet. **28,** 116-123.

Kurihara, N., Menaa, C., Maeda, H., Haile, D.J. and Reddy, S.V. (2001). Osteoclast-stimulating factor interacts with the spinal muscular atrophy gene product to stimulate osteoclast formation. J.Biol.Chem. **276,** 41035-41039.

Ladjali-Mohammedi, K., Grapin-Botton, A., Bonnin, M.A and Le Douarin, N.M. (2001). Distribution of HOX genes in the chicken genome reveals a new segment of conservation between human and chicken. Cytogenet. Cell Genet **92,** 157-161.

Lawler, J., McHenry, K., Duquette, M. and Derek, L. (1995). Characterization of human thrombospondin-4. J. Biol.Chem. **270,** 2809-2814.

Li, M. D. and Kane, J.K. (2003). Effect of nicotine on the expression of leptin and forebrain leptin receptors in the rat. Brain Res. **991,** 222-231.

Li, M.D., Kane, J.K. and Konu, O. (2003). Nicotine, body weight and potential implications in the treatment of obesity. Curr. Top.Med.Chem **3,** 899-919.

Michel, U., Farnworth, P. and Findlay, J.K. (1993). Follistatins: more than follicle stimulating hormone suppressing proteins. Mol. Cell Endocrinol. **91,** 1-11.

Nanda, I., Sick C., Munster, U., Kaspers, B., Schartl, M., Staeheli, P. and Schmid, M. (1998). Sex chromosome linkage of chicken and duck type I interferon genes: further evidence of evolutionary conservation of the Z chromosome in birds. Chromosoma **107,** 204-210.

Nanda, I., Zend-Ajusch, E., Shan, Z., Grutzner, F., Schartl, M., Burt, D.W., Koehler, M., Fowler, V.M., Goodwin, G., Schneide,r W.J., Mizuno, S., Dechant, G., Haaf, T and Schmid, M. (2000). Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene DMRT1: a comparative review on avian sex determination. Cytogenetics and Cell Genetics 89**,** 67-78.

Nanda, I., Haaf, T., Schartl, M., Schmid, M and Burt, D.W. (2002). Comparative mapping of Z-orthologous genes in vertebrates: implications for evolution of avian sex chromosomes. Cytogenet. Genome Res. **99,** 178-184.

Newton, G., Weremowicz, S., Morton, C.C., Jenkin, N.A., Gilbert, D.J., Copeland, N.G. and Lawler, J. (1999). The thrompospondin-4 gene. Mamm. Genome **10,** 1010-1016.

O'Brien, S. J. and Nash, W.G. (1982). Genetic mapping in mammals: Chromosome map of domestic cat. Science **216,** 257-265.

Pedersen, S.B., Kristensen, K. and Richelsen, B. (2003). Anti-glucocorticoid effects of progesterone in vivo on rat adipose tissue metabolism. Steroids **68,** 543-550.

Pertsemlidis A. and Fondon J.W. III (2001). Having a BLAST with bioinformatics (and avoiding BLASTphemy). *Genome Biology* **2,** 2002.1-2002.10.

Rapp, J. P., Wang, S. M. and Dene, H. (1989). A genetic polymorphism in the renin gene of dahl rats co-segregates with blood pressure. Science **243,** 542-44.

Rescher, U. and Gerke, V. (2004). Annexins-unique membrane binding proteins with diverse functions. J.Cell Sci. **117,** 2631-639.

Sasaki, O., Odawara, S., Takahashi, H., Nirasawa, ,K., Oyadama, Y.,Yamamoto, R., Ishii, K., Nagamine, Y., Takeda, H., Kobayashi, E. and Furukawa, T. (2004). Genetic mapping of quantitative trait loci affecting body weight, egg character and egg production in F2 intercross chickens. Anim. Genet. **35,** 188-194.

Sattler, T and Furll, M. (2004). Creatine kinase and aspartate aminotransferase in cows as indicators for endometritis. J. Vet. Med A Physio. Pathol. Clin. Med. **51,** 132-137.

Smith, J., Paton I.R., Horvat, S., Medrano, J.F and Burt, D.W. (2000). Mapping the RAIDD gene of chicken (Gallus Gallus): Identification of a region homologous to the mouse high growth region. Mamm. Genome **11,** 706-9.

Suchyta, S.P., Cheng, H H., Burnside, J and Dodgson, J.B. (2001). Comparative mapping of chicken anchor loci orthologous to genes on human chromosomes 1, 4 and 9. Anim. Genet. **32,** 12-18.

Sztalryd, C., Hamilton, J., Horowitz, B.A., Johnson, P. and Kraemer, F.B. (1996). Alterations of lipolysis and lipoprotein lipase in chronically nicotine-treated rats. Am. J. Physiol. Endocrinol. Metab. **270,** E215-23.

Tucker, R.P., Adams, J.C. and Lawler, J. (1995). Thrombospondin-4 is expressed by early osteogenic tissues in the chick embryo. Dev.Dyn. **203,** 477-90.

Tuiskula-Haavisto, M., Honkatukia, M., Vilkki J., deKoning, D.J., Schulman, N.F and Maki-Tanila, A. (2002). Mapping of QTL affecting quality and production traits in egg layers. Poultry Science **81,** 919-27.

Vanhoutteghem, A and Djian, P. (2004). Basonuclin 2: an extremely conserved homolog of the zinc finger protein basonuclin. Proc. Natl. Acad.Sci. **101,** 3468-3473.

Van Kaam, J.B.C.H.M., Van Arendonk, J.A.M., Groenen, M.A.M., Bovenhius, H., Vereijken, A.L.J.,Crooijmans, R.P.M.A., Van der Poel, J.J. and Veenendaal, A.(1998). Whole genome scan for quantitative trait loci affecting body weight in chickens using a three generation design. Livest.Prod.Sci.**54,** 133-150.

Volpini, L.M., Calnek, B.W., Sneath, B., Sekellick, M.J. and Marcus, P.I. (1996). Interferon modulation of Marek's disease virus genome expression in chicken cell lines. Avian Dis. **40,** 78-87.

Yona, S., Ward, B., Buckingham, J.C., Perretti, M and Flower, R.J. (2005). Macrophage biology in the Anx-A1-/- mouse. Prostaglandins. Leukot. Essent. Fatty acids **72,** 95-103.

Zhou, H., Li, H. and Lamont S.J. (2003). Genetic markers associated with antibody response kinetics in adult chickens. Poultry Science **82,** 699-708.

Website References:

1. http://www.genome.iastate.edu/chickmap

2. https://acedb.asg.wur.nl/

3. http://www.ensembl.org

4. http://www.ncbi.nlm.nih/gov

Table 1 Probable quantitative trait loci locations on chicken GGAZ and their flanking markers

Trait	Location[1] (bp)	Location[2](cM)	Flanking markers	Reference
Growth (112-200d)	1,880,642	22	MCW0055	Kerje et al., 2003
Body weight (200d)	11,070,329	22	ADL0273	Kerje et al., 2003
Age at first egg (day)	14,783,516-15,457,996	28	ADL0201-MCW0241	Sasaki et al., 2004
Antibody response to *Brucella abortus*	14,783,516-17,635,971	28	ADL0201-ADL0250	Zhou et al., 2003
Egg shell thickness	16,538,085	36	LEI0229	Sasaki et al., 2004
Egg shell strength	16,770,587-19,749,500	47	MCW0154-LEI0254	Sasaki et al., 2004
Age at first egg (day)	5,209,429-15,297,279	63-104	MCW258-MCW246	Tuiskula-Haavisto et al., 2002
Egg weight (40 wk)	5,209,429-15,297,279	63-104	MCW258-MCW246	Tuiskula-Haavisto et al., 2002
Egg weight (40-60 wk)	5,209,429-15,297,279	63-104	MCW258-MCW246	Tuiskula-Haavisto et al., 2002
Egg shell strength (40 wk)	5,209,429-15,297,279	63-104	MCW258-MCW246	Tuiskula-Haavisto et al., 2002
Egg number (41- 60 wk)	5,209,429-15,297,279	63-104	MCW258-MCW246	Tuiskula-Haavisto et al., 2002
Body weight (239 d)	UN[3]	96	LEI0075-LEI0123	Sasaki et al., 2004
Abdominal fat weight	22,950,349-31,399,653	127	LEI0111-LEI0075	Ikeobi et al., 2002
Abdominal fatness	22,950,349-31,399,653	127	LEI0111-LEI0075	Ikeobi et al., 2002

[1]Sequence map
[2]Genetic map
[3]Unassigned

Table 2 Orthologous comparison of probable quantitative trait loci locations on GGAZ with mouse and human genomes

Trait	Location		
	GGA	MMU	HSA
Growth	22	13D2.1	5q11.2
Body weight	22	13D1	5q11.2-q14.3
Age at first egg	28	4C3	9p23-p21.3
Antibody response to *Brucella Abortus*	28	19C1	9p21.3
Egg shell thickness	36	NH	NH
Egg shell strength	47	19B	9q21.13
Age at first egg(day)	63-104	4C3-C4	5p12
Egg weight (40 wk)	63-104	4C3-C4	5p12
Egg weight (40-60 wk)	63-104	4C3-C4	5p12
Egg shell strength	63-104	4C3-C4	5p12
Egg number (18-40 wk)	63-104	4C3-C4	5p12
Egg number (41-60 wk)	63-104	4C3-C4	5p12
Body weight (239d)	96	NH	NH
Abdominal fat weight	127	4B3	5q14.3
Abdominal fatness	127	4B3	5q14.3

NH=No homology

Table 3 Putative candidate genes around QTL locations on chicken GGAZ and their respective homologues in mouse and human

Trait	Probable Location	Putative Candidate Genes		
		GGA	MMU	HSA
Growth Body weight	22	Claustrin	cAMP-specific 3',5'-cyclic phosphodiesterase 4D Transportin 1 Mitochondrial ribosomal protein S27 Microtubule-associated preotein 1B scamp 1	Serine/threonine-protein kinase PLK2 Ras-related protein Mitochondrial ribosomal protein S2 Microtubule-associated preotein 1B Pentatricopeptide repeat domain 2 Zinc finger protein 266
Antibody response to *Brucella abortus*	28	Proprotein convertase PC6 nuclear orphan receptor ROR-β tight junction protein Annexin A1 (Annexin I) (Lipocortin I)	Annxexin A1 (lipocortin 1) RNP particle component Transmembrane cochlear-expressed protein 1 Aldehyde dehydrogenase 1A1 Zinc finger protein 216 Guanine deaminase	Long transient receptor potential channel 3 transmembrane protein 2 annexin A1
Egg shell Strength	47	Annexin A1 (Annexin I) (Lipocortin I) nuclear orphan receptor ROR-β bZIP protein E4BP4 Proprotein convertase PC6 transducin-like enhancer of split 4 40S ribosomal protein S15 Trypsin II-P29 precursor	Endometrial progesterone-induced protein Guanine nucleotide-binding protein, alpha-14 subunit Riboflavin kinase IP63 protein Proprotein convertase subtilisin/kexin type 5 precursor Chorea-acanthocytosis homolog Guanine nucleotide-binding protein G (q), alpha subunit. β-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase Forkhead box protein B2 transient receptor potential cation channel, subfamily M, member 6	osteoclast stimulating factor 1 Vacuolar protein sorting 13A (Chorein) proprotein convertase subtilisin/kexin type 5 Guanine nucleotide-binding protein, alpha-14 subunit Riboflavin kinase guanine nucleotide binding protein (G protein) (β-1,6-N-acetylglucosaminyltransferase) phosphoserine aminotransferase 1

Egg weight (40 wk)	63-104	Centromere protein H	multiple PDZ domain protein	Growth hormone receptor precursor (GH receptor)
Egg weight (40-60 wk)	63-104	Thrombospondin-4	RNP particle component	(Somatotropin receptor)
Egg shell strength (40wk)	63-104	fibroblast growth factor 10	Nuclear factor 1 B-type	Selenoprotein P precursor
Egg number (18-40 wk)	63-104	ZOV3 gene product	Zinc finger DHHC domain containing protein 21	Small inducible cytokine A28 precursor
Egg number (41-60 wk)	63-104	Laminin and collagen receptor	cerberus 1 homolog	Integrin α-1 (Laminin and collagen receptor
Age at first egg (day)	63-104	NADH dehydrogenase	small nuclear RNA activating complex, polypeptide 3	Hydroxymethylglutaryl-CoA synthase, cytoplasmic
		3-hydroxy-3-methylglutaryl-CoA reductase	lens epithelium-derived growth factor;	NAD(P) transhydrogenase, mitochondrial precursor
		claustrin	Weakly similar to hypothetical 71.7 kDa protein	Annexin II receptor
		Sodium/potassium/calcium exchanger 2 precursor	basonuclin 2	Collagen receptor
		endophilin	Adipophilin (Adipose differentiation-related protein)	Fibroblast growth factor-10 precursor
		Growth hormone receptor precursor	ADAMTS-like protein 1 precursor (Punctin)	Programmed cell death protein 9
		Hydroxymethylglutaryl-CoA synthase, cytoplasmic	solute carrier family 24 (sodium/potassium/calcium exchanger)	Potassium/sodium hyperpolarization-activated cyclic –nucleotide-gated channel 1
		p52 pro-apototic protein	Ras-related GTP-binding protein	Zinc finger protein 131
		Insulin protein ISL-1 gene enhancer	40S ribosomal protein S6 (Phosphoprotein NP33).	gene similar to embigin
		Follistatin precursor	Cancer related gene-liver 1	Insulin gene enhancer protein ISL-1
		Spindling	Myeloid/lymphoid or mixed lineage – leukemia	poly (ADP-ribose) polymerase family
		polymerase (DNA directed) kappa	Translocation homolog to 3 homolog	Molybdenum cofactor synthesis protein 2 large subunit
		Glycine dehydrogenase [decarboxylating], mitochondrial precursor	Interferon β precursor	Myosin tail domain containing protein
		40S ribosomal protein S6	MKIAA1797 protein	Small nuclear RNA activating complex, polypeptide
		bZIP protein E4BP4	α-interferon	PC4 and SFRS1 interacting protein 1
			interferon α 14	
			interferon α family, gene 12; interferon α 12	
			interferon α family, gene 13; interferon α 13;	
			interferon α 6T	
			limitin	
			interferon α family, gene 11	
			Interferon α-1 precursor	
			interferon tau-1	
			Interferon α-4 precursor	
			S-methyl-5-thioadenosine phosphorylase	

			Cyclin-dependent kinase 4 inhibitor A Cyclin-dependent kinase 4 inhibitor B zinc finger protein 352 L1Md-A13 repetitive sequence RNP particle component (Fragment) 5,6-dihydroxyindole-2-carboxylic acid oxidase precursor	
Abdominal fat weight	127	ORF2 protein Limb expression 1 mature protein acetylcholine receptor protein putative alpha3 fucotransferase pro--neuregulin 1 precursor (contains neuregulin which has- acetylcholine receptor inducing activity) progesterone receptor binding protein phosphodiesterase 6 β subunit purpurin precursor neuronalacetylcholine receptor ATP binding fructose biphosphate aldolase B cassette creatine kinase, sarcomeric mitochondrial precursor dihydrofolate reductase cartilage link protein	RNP particle component (Fragment) olfactory receptor 273 T-cell acute lymphocytic leukemia-2 protein homolog olfactory receptor 270 cystatin and DUF19 domain-containing protein 1 olfactory receptor 275 Fukutin olfactory receptor 272 Snap3B protein ATP-binding cassette UV excision repair protein RAD23 homolog B zinc finger protein 462 Kruppel-like factor 4 inhibitor of kappa light polypeptide enhancer in B-cells catenin α-like 1 gene similar to brain protein Protein tyrosine phosphatase Glyceraldehyde 3-phosphate dehydrogenase A-kinase anchor protein 2 polydomain protein muscle, skeletal, receptor tyrosine kinase	Adapter-related protein complex 3 beta 1 subunit lipoma HMGIC fusion partner-like 2 junction-mediating and regulatory protein Single-stranded DNA-binding protein 2 Thrombospondin 4 precursor cardiomyopathy associated Secretory carrier-associated membrane protein 1 Dimethylglycine dehydrogenase, mitochondrial precursor AASA9217 membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1 Hyaluronan and proteoglycan link protein 1 precursor Homer protein homolog 1 Arylsulfatase B precursor developmentally regulated protein TPO1 Zinc finger CCHC domain containing protein 9 cytosolic phosphoprotein DP58 DNA-repair protein XRCC4 (X-ray repair cromethyltransferase 2 cross-complementing protein 4 PAP associated domain containing 4

	Dihydrofolate reductase
	Ras protein-specific guanine nucleotide-releasing factor 2
	Creatine kinase, sarcomeric mitochondrial precursor
	Cytoplasmic acetyl-CoA hydrolase 1
	APG10 autophagy 10-like protein
	ribosomal protein S23
	DNA mismatch repair protein Ms
	Versican core protein precursor (Large fibroblast proteoglycan)
	Developmentally regulated endothelial cell locus 1 protein

CHAPTER 4

QUANTITATIVE TRAIT LOCI ANALYSIS FOR GROWTH AND SKELETAL TRAITS IN HIGH AND LOW GROWTH CHICKENS REVEALS SEX SPECIFICITY[1]

[1] Georgina Ankra-Badu, Samuel Aggrey et al., to be submitted to Genome Biology

ABSTRACT

Traits of economic importance are affected by several gene loci together with environmental factors. Trait genes can however be detected by the identification of quantitative trait loci (QTL) that underlie these traits. An F_2 population was established from a chicken line selected for high (HL) or low (LH) growth. We generated 695 F_2 individuals from reciprocal F_1 crosses that were used to localize QTL for age-related body weight (BW), shank length (SL) and shank diameter (SD). QTL mapping revealed 11 BW QTL that additively explain 64% of the phenotypic variance. Most of the BW QTL individually explained 2-4% of the phenotypic variance. However, a pleiotropic QTL on GGA4 explained 7-11% of the variance and affected BW at 5 to 9 weeks, SL and SD at 9 weeks. The QTL on GGA4 had the strongest effect on SL and SD, and explained 18% and 21%, respectively of the phenotypic variance. A male-specific BW QTL on GGA3 at 179 cM had no effect in females, but a genome-wide effect on males. Many QTL account for BW and growth and explains at least in part, the continuous success selection for growth has achieved in chickens for more than 50 years despite intense selection. Orthologous comparison of QTL regions with mouse and human genomes revealed several candidate genes for the study of genetic architecture of growth and skeletal development. Since the foundation population was established with commercial broiler strains, it is possible that QTL identified from this study are still segregating.

INTRODUCTION

Traits that exhibit Mendelian inheritance are influenced by mutations in single or few genes that determine the phenotype. However, most traits of economic importance

follow a complex mode of inheritance, and are affected by several gene loci, gene interactions and environmental factors. In addition, most economically important traits have negative genetic correlation with fitness therefore simultaneous improvement of these traits is challenging when traditional selection methods are utilized. The use of molecular markers that are directly or indirectly linked to quantitative trait loci (QTL) could increase the response of these traits to selection without a corresponding decrease in fitness [1, 2]. Identification of QTL regions for traits will also allow for the delineation of individual genes that may underlie such traits.

Growth related traits are of particular importance to the meat-type chicken industry, and as a result several studies have been conducted to locate body weight (BW) and growth QTL in various experimental populations. Various BW QTL were identified on GGA1, 2, 4, 7, 8 and 13 from a broiler-layer cross [1]. Significant BW QTL have also been identified on GGA2, 4 and 27 [3, 4]. The growth QTL on GGA2 and 4 explained a relatively large proportion of the phenotypic variance in these populations the studies were conducted with, suggesting that GGA2 and GGA4 may contain several genes that are associated with variations in growth. There are also reports on many QTL with minor additive effects associated with growth in divergently selected White Plymouth Rock chickens [5].

Body weight is a quantitative trait consisting of the weights of lean and fat compounds, bones and body fluids [6]. Studies on mice and chickens suggest that genetic regulation of body weight varies with age [7-10]. Early growth in poultry is more flexible than later growth, and genes affecting the developmental period may be different from genes operating at the maturation period of growth [11]. There are also separate sets of

QTL for early and late growth in mice [8]. It has also been shown that while additive effects on growth were more pronounced at a later age epistasis played a more important role in growth (1-8 weeks) in layer-broiler and White Plymouth Rock crosses [9, 10]. There is no significant genetic correlation between hatch weight and later BW [12]. These studies suggest the necessity to delineate the genetic factors affecting BW during early and late growth.

Remarkable improvements have been made in selection for BW using mass selection [13]. This achievement has been accompanied by several latent "pathologies" such as leg and skeletal disorders, excessive body fat and possible reduction in overall fitness [14, 15]. Longer shanks have been found to contribute to leg abnormalities in chicken, and as a result meat-type chickens are selected for shanks that are short relative to their BW [2]. It is expected that few genes control both shank length (SL) and shank diameter (SD) since heritability for both traits are high [16]. Two shank weight QTL have been identified on GGA1 and 27 that explained about 20% of the phenotypic variation in divergent population selected for growth [17]. The importance of identifying QTL underlying growth and skeletal strength in the same population cannot be understated. Since SL and SD are correlated with growth, identification of genes that control shank length will aid selection for skeletal integrity in fast growing chickens.

We have generated a large intercross population from a chicken population divergently selected for BW at 8 weeks of age [7]. Whereas BW QTL reported to date come from either an intercross within a breed [5, 18] or between breeds [1, 3, 19, 20], the divergent lines used for the current study was initiated from a base population that was constituted from commercial poultry strains. The growth and physiological characteristics

of the divergent lines are well documented [21-26]. A gene by age interaction from a global gene expression study on the divergent lines has been reported, further suggesting the possibility of different genes affecting growth at different ages [27].

The objectives of the current study were to identify age-related BW, SL, and SD QTL in a chicken population divergently selected for high or low BW and to identify positional candidate genes in these QTL regions by orthologous comparative mapping.

MATERIALS AND METHODS

Experimental animals

The high (HG) and low (LG) growing lines were developed in a chicken population divergently selected for 14 generations for 8 and 36 weeks BW [7]. A reciprocal F_1 cross was generated from the base population by mating 5 HG males to 16 LG females and 5 LG males to 9 HG females. From the F_1 generation, 3 HL (HG sire x LG dam) males were intercrossed with 30 HL females and 2 LH (LG sire x HG dam) males were intercrossed with 20 LH females to develop the F_2 population. Six hundred and ninety five F_2 chickens consisting of 50 full-sib families were used for the QTL analysis. These birds were raised in four hatches under standard management practices for 9 weeks and were fed a standard broiler diet of 3050Kcal ME from 0-3 weeks and 3100 Kcal ME from 4-9 weeks *ad libitum*. Traits measured in the F_2 population were BW at 0, 1, 3, 5, 7 and 9 weeks, SL and SD at 9 weeks.

Genotyping

A 100 µl sample of blood was obtained from each bird and DNA was extracted by phenol-chloroform extraction. Microsatellite markers used in the analysis were chosen

from the poultry consensus map [47] based on their location on the chromosome and informativeness on each F_1 sire family. Genotyping of the DNA samples was done based on one hundred and nine markers from 20 autosomal linkage groups. Depending on their size and amplification conditions some markers were combined (2~10) for multiplex PCR amplification. Fluorescent microsatellite analysis was done on an ABI 3700 DNA sequencer (Applied Biosystems, Foster City, CA, USA). Each genotype was interpreted using both the GeneScan Analysis 3.7 and Genotyper Analysis 3.7 softwares (Applied Biosystems, Foster City, CA, USA). The GEMMA database was used to manage the informativity tests [48].

QTL analysis and comparative mapping

An interval mapping method of analysis [49] was conducted using the QTL express software program [50]. The model for the analysis included hatch, family and sex as fixed effects. A one-QTL model was applied by searching at 1-cM intervals along the chromosome and was assessed by comparing the F-ratio of a model with one QTL against a model with no QTL. Additive and dominance effects together with QTL by sex interaction were estimated by least squares for each putative QTL for all traits. A separate analysis was conducted for each sex in order to identify sex specific QTL. Probabilities for significant and suggestive linkage were simulated by randomization using 1000 permutations of the data [51]. Genome wide analysis was conducted to determine the F-ratio thresholds at which there were 1% and 5% probability of locating a QTL on the genome where no QTL was fitted [52]. The highest F-ratio was selected and test statistic thresholds at 1% and 5% were approximately 6.5 and 5.5 respectively. Tests for chromosome-wide significance were also conducted and a confidence interval for the

position of each QTL was obtained by bootstrapping with a 1000 samples. The phenotypic variance explained by significant QTL was also calculated for each trait and was estimated as the difference in percent ratio between the full and reduced sum of squares. In order to identify putative positional candidate genes in QTL regions across species, orthologous comparative mapping was done by blasting the primer sequence of the markers encompassing identified QTL against the draft chicken sequence in the Sanger Institute website (www.ensembl.org) using the basic local alignment search tool (BLASTN). The selected sequences were compared with the mouse and human genomes to identify regions of homology. Information on genes located close to or at QTL regions together with their homologs in humans and mice was obtained from www.ensembl.org.

RESULTS

This study was conducted by analyzing seven age-related BW traits, SL and SD on 20 autosomes. Table 1 shows growth performance and bone parameters according to sex. Table 2 shows the total number of markers used in the analysis, their respective map lengths, and the first and last markers for each chromosome, and Table 3 summarizes suggestive and significant QTL regions, their flanking markers and respective confidence intervals. Twenty regions on twelve chromosomes exceeded the 5% chromosome-wide significant threshold while 4 regions exceeded the 1% and 5% genome-wide thresholds on GGA2, 3, 4 and 5.

A suggestive BW0 QTL was found on GGA3. Even though the heritability of BW0 is moderately high, maternal influences associated with this trait are also high [28]. Therefore, hatch weight reflects the maternal phenotype more than the genotype of the

progeny. However, reciprocal F_1 crosses were performed to reduce the influence of maternal effect on the traits studied as evidenced from BW0 from the reciprocal crosses (Table 1). There was no BW QTL at 1 week, and only one suggestive BW QTL at 3 weeks of age on GGA4. There were two BW QTL at 5 weeks on GGA2 and GGA4, respectively. Body weight QTL at week 5 located on GGA2 at 73 cM was significant at a 1% genome-wide threshold level. The region on GGA4 enclosed by *MCW0240* and *LEI0073* (200-218cM) harbors QTL for BW at 5, 7 and 9 weeks of age. Test statistic for the BW QTL on GGA4 exceeded the 1% threshold level on the genome-wide level. Chromosome-wide BW QTL at various ages and QTL for SL and SD were also found on GGA1, 2, 3, 4, 5, 13 and 26. Genome-wide significant QTL were found for *Growth3 and Growth5* on GGA4, and for *Growth4* on GGA3, 4 and 5. Genome-wide highly significant QTL were also found for SL and SD in the same region of the BW QTL on GGA4.

The means and standard errors for additive and dominance effects for both sexes as well as the proportion of the phenotypic variance explained by the QTL are shown in Table 4. The QTL effects ranged from 2% to 21% with GGA4 accounting for the largest proportion of the phenotypic variance. The 16 BW QTL and 22 *Growth* QTL explained 64% and 69% of the phenotypic variances in the F_2 population respectively. Each BW QTL on GGA1, 2, 3, 5, 13 and 26 explained only a small proportion of the phenotypic variance, 2-3% in the F_2 population. The BW QTL on GGA4 explained 7-11% of the phenotypic variance. The *Growth* QTL explained 2-9% of the phenotypic variance. The QTL on GGA4 had the strongest effect on SL and SD, and they explained 18% and 21%, respectively of the phenotypic variance. Three SL QTL on GGA1, 2 and 4 cumulatively

explained 24% of the phenotypic variance and three SD QTL on GGA3, 4, and 26 explained 27% of the phenotypic variance.

The BW7 QTL on GGA3 showed sexual dimorphism when the data was analyzed with sex interactions (additive and dominance) included in the model (Table 4). There was a 3 fold difference in additive effect between males and females for the BW9 QTL. When the data was analyzed separately for each sex, it became apparent that the BW7 and *Growth4* QTL on GGA3 are male specific (Table5; Figure 1). The BW7 and *Growth4* QTL on GGA3 were not significant in the female, but significant on the genome-wide level in the males.

DISCUSSION

This study revealed 11 genome- and chromosome-wide significant age-related BW QTL that additively explains 64% of the phenotypic variance in the F_2 population. Most of the BW QTL individually explained a small proportion (2-4%) of the phenotypic variance. Most of the BW QTL revealed from this study had minor individual effects, which was consistent with reports from other studies [1, 3, 5, 19], however, the BW QTL on GGA4 explained 7-11% of the phenotypic variance. The statistical analysis assumes that the alternative QTL alleles are fixed in the divergent lines. Any violations of this assumption would lead to the underestimation of the QTL effect. Therefore, it is possible that the QTL effects are larger than reported, but their real effects will be known when the causative genes that underlie the QTL are identified. Also QTL effects depends on the genetic background of the population(s) investigated and huge phenotypic differences

between two populations do not necessarily imply the existence of QTL with large effects [5].

There was no significant hatch weight QTL at the genome-wide level from our study. Kerje et al. [29] also failed to identify a QTL for hatch weight in an F_2 cross between the white Leghorn and the jungle fowl. Hatch weight as a trait is associated with low additive genetic effect [30] and greatly influenced by the maternal phenotype, most likely through egg and egg weight [28]. Genetic decomposition of factors that affect hatch weight and early growth would delineate the true value of the trait, and that could potentially allow the trait to be mapped. The fact that no QTL was identified in the current study for BW or growth during the early juvenile period (0-3 weeks) is surprising considering the selection scheme that was used. The HG and LG chickens were selected for high or low body weight at 8 weeks and approximately 36 weeks and their growth were significantly diverged during the early juvenile period (0-3 weeks). We identified two QTL for BW at week 5 on GGA2 and 4. It has been reported that growth before 46 days is affected by several QTL thus epistatic interaction is very important during this period [9, 10]. They further estimated that epistasis explained about 70% of the variation in growth from 1 to 8 weeks. The degree to which epistasis affects early growth may be population dependent. The allele frequency in any population depends on the adaptive environment of that population. Therefore, the growth dynamics of a jungle fowl-layer cross for example is very different from that of a broiler population and consequently early or late growth from different crosses are also not comparable. The true nature of what constitutes early and late growth in any population will depend on what constitutes growth: skeletal development, cell growth, lean accretion or cell atrophy, muscle and fat

deposition. Therefore the biological time clock of the population, e.g. point of inflection should warrant more attention than an arbitrary chronological age. In most instances the additive and dominance effects were in opposite directions. The standard errors associated with most dominance estimates were too high and thus reduce the confidence in the dominance estimates. We did not test for epistasis in this study nevertheless we still believe that both dominance and epistasis have affects on the QTL studied. Chromosomes 2 and 4 appear to be very critical regions for the regulation of several traits in chickens regardless of population type. Significant QTL with very large effects on BW have been located on GGA4 [3, 19, 29]. Several important QTL for egg traits and BW have been located on GGA2 and GGA4, with QTL on GGA4 having the highest effect [4]. A QTL for BW at 6 weeks has also been found on GGA2 at position 302 cM [1] while the current study locates QTL for BW at 7 and 9 weeks at positions 379 cM and 375 cM, respectively. Other QTL have been identified for BW at 13 and 16 weeks on GGA1 and GGA2, using an F_2 cross from a slow growing Japanese breed and a fast growing Plymouth Rock breed [31]. The QTL in their study explained a very large proportion of the phenotypic variance (12.2-41.7%). Significant loci for BW in our study were few and were only associated with GGA2 and GGA4. Artificial selection reduces differences in allele frequency for growth QTL thus the number of samples used in QTL studies of this kind of population must be increased to detect QTL effects [32].

The age-specific trend in growth was evident in the variations in additive effects for the different stages of growth. Generally, additive effects for growth and the proportion of the phenotypic variance explained by the QTL increased with age for both males and females. The proportion of the phenotypic variation explained by the QTL

increased at later ages due to the higher number of identified QTL and the higher additive effects [33]. It is obvious that the growth QTL on GGA4 only affected BW from 5 weeks onwards and its effect increased with age.

It is well documented that male chickens grow faster and are heavier than female chickens [11, 25] Therefore it is not unusual to expect some QTL to exhibit sexual dimorphism. We tested each QTL for sexual dimorphism in this study and observed that BW9 QTL on GGA3 at 179 cM had more positive influence on males than females. The biological mechanisms underlying sexual dimorphism for this QTL are unknown, but influence of sex hormones cannot be ruled out. The results from the present study indicate that it is important to differentiate between the sexes in a QTL analysis, since a male-specific BW7 QTL was identified on GGA3. Other sex-specific QTL have been reported [34-37], but the genetic mechanisms that underlie such sex-specific QTL are still unknown. It is important to ascertain whether a QTL is sex-specific, sex-influenced or neither before any fine mapping strategies are pursued. It will save the number of individuals selected to be genotyped. Additionally the expected genetic progress associated with marker-assisted selection or gene introgression using such sex-specific and sex-influenced QTL can be estimated without bias.

Shank length and SD accounted for the highest proportion of the variance explained by the QTL (18% and 21%, respectively). This result is remarkable but not surprising since SL is highly heritable. Due to the heavy emphasis on improvement in growth and the subsequent negative effect on fitness, broiler chickens usually develop leg problems. Heavy chickens with long shanks also exhibit signs of leg disorders [2] thus broiler chickens are selected for shorter shank length [38]. In turkeys selection for

increased shank width also increased shank weight and the weights of the tibia and femur leading to an increase in leg bone development [39]. Body weight also increased with an increase in shank width. This makes the QTL on GGA4 very important because the QTL allele originates from the high line and has strong positive effects on both growth (5-9 weeks) and SD. Considering the fact the heritability of SD, SL are high, and that of BW and growth are moderate to high, and genetic correlations among age-related BW, growth, SD and SL are also high, it is possible that GGA4 harbors a major gene or few genes with large pleoitropic effects and the aforementioned traits. Our high and low growing lines provide a good model for growth and skeletal development in both humans and poultry and to identify already localized genes that can be potentially considered as candidate genes. The additive effect of each of the significant QTL were positive indicating that the HG alleles were responsible for increasing values of the BW, SL and SD. Body weight and growth are under the influence of a large number of QTL, in contrast to shank parameters. The fact that many QTL account for BW and growth most likely explains at least in part, the continuous success that quantitative selection for growth has achieved in chickens for more than 50 years despite intense selection.

One of the major goals in QTL research is ultimately identify the genes underlying the QTL and the regulatory mutations in the gene(s) causing the QTL effect. The confidence interval for the probable QTL location is usually large nevertheless genes located within the probable QTL regions will be the logical next step to consider as potential candidates. We used the microsatellite loci flanking the estimated QTL location on the draft chicken sequence (www.ensemble.org) to identify candidate genes in the QTL region. The region encompassing *LEI0147* and *MCW0096* on GGA2 has thirty-three

genes covering about 7.69 MB (7,688,912 bp). Orthologous comparison of this region with the mouse and human genomes yielded 46 additional candidate genes. Extracting potential candidate genes at probable QTL regions through orthologous comparison has been used successfully for GGAZ [40]. Among the potential candidate genes for the BW QTL on GGA2 are laminin α 1 chain, myosin regulatory chain, myomesin, colon cancer associated protein, oxysterol binding protein related protein 1 (*OSBP*), and myo-inositol monophosphatase.

Laminins are complex glycoproteins that bind to cells through a high affinity receptor. It is a cellular component and is a major component of basement membranes (GO 0005604). It mediates the attachment, migration and organization of cells into tissues during embryonic development by interacting with other extracellular components. Laminins influence signal transmissions during muscle formation and regeneration and mutations in laminins have been found to affect muscle formation [41]. Laminin α 1 has been shown to improve mobility and lifespan in mice suffering from muscular dystrophy [42]. The ankyrin repeat protein (*ARP*) that is involved in muscle formation was located in the BW QTL region on GGA2. Messenger RNA of *ARP* has been found in the myotubes during muscle formation. Oxysterol binding proteins (*OSBPs*) are involved in the molecular functions and biological processes of phospholipid binding and cholesterol metabolism respectively. The *OSBP*s contain ankyrin repeats that are associated with muscle formation. Oxysterol binding proteins are also involved in cholesterol homeostasis, calcium uptake and cell differentiation [43, 44]. Aquaporin 4(*AQP4*) is the candidate gene identified through homologous comparison of the BW QTL region on GGA4 and the mouse and human genomes. The *AQP4* gene is

involved in biological processes and molecular functions of transporter activity. Interestingly, *AQP4* encodes a protein that contains ankyrin repeats that are involved in muscle formation. *AQP4* is enriched at the sarcolemma of skeletal muscle and is lost after the onset of muscle degeneration [45]. The *MCW0240* and *LEI0073* loci flanking the BW and shank width QTL on GGA4 presently harbors 35 genes spanning about 11.76 Mb (11,760,055 bp). This region contains the neuronal acetylcholine receptor (*nAchR*), FGF binding protein (*FGFBP*) and cholecystokinin type A receptor (*CCKAR*) genes among others. The neuronal acetylcholine receptor is found in the neuromuscular junction and interacts with nicotine to regulate appetite, body fat and weight gain in rats [46].

Any inferences from positional candidate genes should be done with caution since the confidence interval associated with the QTL are generally large. Moreover, the number of potential candidate genes in probable QTL regions from a gene mapping studies can be overwhelming. Identification of the causative genes to the traits studied will require fine-scaled mapping of the QTL regions and differentially expressed genes within refined QTL regions in the founder lines. For example, the myo-inositol monophosphatase gene located within the BW QTL on GGA2 is differentially expressed between the divergent lines used in this study (Cogburn et al., unpublished data).

SUMMARY

Our study has identified 2 major and 9 minor QTL for growth and SD on GGA2 and GGA4. The BW QTL identified on GGA4 has strong additive and pleiotropic effect as it affects BW from 5-9 weeks of age, as well as SL and SD at 9 weeks. The additive

QTL effect of BW at 9 weeks is about 85 g; replacing the SGL allele with the alternative HG allele will increase BW by about 170 g. Since the foundation population of the divergent lines was established with commercial strains, it is possible that QTL identified from this study are still segregating in commercial broiler strains. The genes within the vicinity of QTL identified provide candidates for further study on the genetic regulation of growth and skeletal traits.

REFERENCES

1. Sewalem A, Morrice DM, Law A, Windsor D, Haley CS: Ikeobi CON, Burt DW, Hocking PM: **Mapping of quantitative trait loci for body weight at three, six and nine weeks of age in a broiler layer cross.** Poultry Sci 2002, **81:** 1775-1781.

2. Deeb L, Lamont SJ: **Genetic architecture of growth and body composition in unique chicken populations.** J Hered 2002, **93:** 107-118.

3. Sasaki O, Odawara S, Takahasi H, Nisawara K, Oyamada Y, Yamamoto R, Ishii K, Nagamine Y, Takeda H, Kobayashi E, Furukawa T: **Genetic mapping of quantitative trait loci affecting body weight, egg character and egg production in F2 intercross chickens.** Anim Genet 2004, **35:** 188-194.

4. Schreiweis M.A, Hester PY, Settar P, Moody DE: **Identification of quantitative trait loci associated with egg quality, egg production, and body weight in an F2 resource population of chickens.** Anim Genet 2006, **37:**106-117.

5. Jacobsson L, Park H-P, Wahlberg P, Fredriksson R, Prez-Encisco M, Siegel PB, Andersson L: **Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens.** Genet Res 2005, **86:** 115-125.

6. Brockmann GA, Haley CS, Renne U, Knott SA, Schwerin M: **Quantitative trait loci affecting body weight and fatness from a mouse line selected for extreme high growth.** Genetics 1998, **150:** 369-381.

7. Ricard F.H: **Essai de selection sur la forme de la courbe de croissance chez le Poulet. Dispositif experimental et premiers resultants.** Ann Genet Sel Anim 1975, **7:** 427-443.

8. Cheverud JM, Routman EJ, Duarte FAM, Van Swinderen B, Cothran K, Peren C: **Quantitative trait loci for murine growth.** Genetics 1996, **142:**1305-1319.

9. Carlborg Ö, Kerje S, Schutz K, Jacobsson L, Jensen P, Andersson L: **A global research reveals epistatic interaction between QTL for early growth in the chicken.** Genome Res 2003, **13:** 413-421.

10. Carlborg Ö, Jacobsson L, Åhgren P, Siegel P, Andersson L**: Epistasis and the release of genetic variation during long-term selection.** Nat Genet 2006, **38:**418-420.

11. Aggrey SE: **Dynamics of relative growth rate in Japanese quail lines divergently selected for growth and their control.** Growth Develop Aging 2003, **67:** 47-54.

12. Abdellatif MA: **Genetic study of Dandarawy chickens: I. Heritabilities and genetic correlations of body weight and weight gain.** Genet Sel Evol 1989, **21:**81-92.

13. Arthur JA, Albers GA: **Industrial perspective on problems and issues associated with poultry breeding.** In Poultry Genetics Breeding and Biotechnology. Edited by Muir WM, Aggrey SE. Oxon: CABI publishing; 2003: 1-12.

14. Marks H.L: **Long-term selection for body weight in Japanese quail under different environments.** Poultry Sci 1996, **75:** 1198-1203.

15. Whitehead CC, Fleming RH, Julian RJ, Sorensen P: **Skeletal problems associated with selection for increased production.** In Poultry Genetics Breeding and Biotechnology. Edited by Muir WM, Aggrey SE. Oxon: CABI publishing; 2003: 1-12.

16. Ricard FH, Rouvier R: **Etude des mesures de conformation du poulet. V. variabilite phenotypique et genetique des mensurations de carcasse dans une souche de type "Cornish".** Ann Zootech 1968, **17:** 445-458.

17. Park H-B, Jacobsson L, Wahlberg P, Siegel PB, Andersson L: **QTL analysis of body composition and metabolic traits in an intercross between chicken lines divergently selected for growth.** Physiol Genomics 2006, **25:** 216-223.

18. Van Kaam JBC, Groenen MAM Bovehius H, Veenendaal A, Vereiken ALJ, Van Arendon JAM: **Whole genome scan in chickens for quantitative trait loci affecting growth and feed efficiency.** Poultry Sci 1999, **78:** 15-23.

19. Jennen DGJ, Vereijken ALJ, Bovenhius H, Crooijmans RPMA, Veenendaal A, Van der Poel JJ, Groenen MAM: **Detection and localization of quantitative loci affecting fatness in broilers.** Poultry Sci 2004, **83:** 295-301.

20. Deeb N, Lamont SJ: **Use of a novel outbred by inbred F-1 cross to detect genetic markers for growth.** Anim Genet 2003, **34:** 205-212.

21. Remignon H, Lefaucher L, Blum JC, Ricard FH: **Effects of Divergent selection for body weight on three Skeletal-muscle characteristics in the chicken.** Brit Poultry Sci 1994, **35:** 65-76.

22. Duclos MJ, Chevalier B, Remignon H, Ricard FH, Goddard C, Simon J: **Divergent selection for high or low growth rate modifies the response of muscle cells to serum or insulin-like growth factor-I in vitro.** Growth Regulat 1996, **6:** 176-184.

23. Mignon-Grasteau S, Beaumont C, Le Bihan-Duval E, Poivey JP, De Rochambeau H, Ricard FH : **Genetic parameters of growth curve parameters in male and female chickens.** Brit Poultry Sci 1999, **40:** 44-51.

24. Mignon-Grasteau S, Piles M, Varona, L., de Rochambeau H, Poivey JP, Blasco A, Beaumont C: **Genetic analysis of growth curve parameters for male and female chickens resulting from selection on shape of growth curve.** J Anim Sci 2000, **78:** 2515-2524.

25. Mignon-Grasteau S, Beaumont C, Ricard FH: **Genetic analysis of a selection experiment on the growth curve of chickens.** Poultry Sci 2001, **80:** 849-854.

26. Beccavin C, Chevalier B, Cogburn LA, Simon J, Duclos, MJ: **Insulin-like growth factors and body growth in chickens divergently selected for high or low growth rate.** J Endocrinol 2001, **168:** 297-306.

27. Jenkins C, LeBihan-Duval E, Simon J, Aggrey SE, Porter T.E, Cogburn LA, Duclos, MJ. **Analysis of the muscle transcriptome in high and low growth**

chickens during the post-hatch period. http://www.intl-pag.org/14/abstracts/PAG14_P798.html 2006.

28. Aggrey SE, Cheng KM: **Animal-model analysis of genetic (co) variances for growth traits in Japanese quail.** Poultry Sci 1994, **73:** 1822-1828.

29. Kerje S, Carlborg O, Jacobsson L, Schutz K, Hartmann C, Jensen P, Andersson L: **The twofold difference in adult size between the red junglefowl and white leghorn chickens is largely explained by a limited number of QTLs.** Anim Sci 2003, **34:** 264-274.

30. Hartmann C: **Selection for yolk production in laying hens.** Ph.D. thesis. Acta Universitasis Agriculturae Sueciae Agraria 298. Swedish University of Agricultural Sciences; 2002.

31. Tatsuda K, Fujinaka K: **Genetic mapping of the QTL affecting body weight in chickens using an F_2 family.** Brit Poultry Sci 2001, **42:** 333-337.

32. Zhu JJ, Lillehoj, HS, Allen PC, Van Tassell CP, Sonstergard TS, Cheng HH, Pollock D, Sadjadi M, Min W, Emara MG: **Mapping quantitative trait loci associated with resistance to coccidiosis and growth.** Poultry Sci 2003, 82: 9-16.

33. Vaughn KT, Pletscher LS, Andrea P, King-Ellison K, Adams E, Erikson C, Cheverud JM: **Mapping quantitative trait loci for murine growth: a closer look at genetic architecture.** Genet Res, **74:** 313-322.

34. Jerez-Timaure NC, Kearney F, Simpson EB, Eisen E.J, Pomp D: **Characterization of QTL with major effects on fatness and growth on mouse chromosome 2.** Obesity Res 2004, **12:**1408-1420.

35. Lionikas A, Blizard DA, Vandenbergh DJ, Glover MG, Stout JT, Vogler GP, McClearn GE, Larsson L: **Genetic architecture on fast-and slow- growth twitch skeletal muscle weight in 200-day-old mice of the C57BL/6J and DBA/2J lineage.** Physiol.Genomics 2003, **16:**141-152.

36. Nuzhdin SV, Pasyukova G, Dilda CL, Zhao-Bang Z, Mackay TFC: **Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster.** Proc. Natl. Acad. Sci 1997, **94:**9734-9739.

37. Abasht B, Pitel F, Lagarrigue S, Le Bihan-Duval E, Le Roy P, Demeure O, Simon J, Cogburn L, Aggrey SE, Vignal A, Douaire M: **Fatness QTL on chicken chromosome 5 and interaction with sex.** Genet. Sel.Evol 2006, **38:** 297-311.

38. Chambers JR: **Genetics of growth and meat production in chickens.** In Poultry Breeding and Genetics. Edited by Crawford RD. Amsterdam: Elsevier; 1990: 599-643.

39. Emmerson DA, Anthony NB, Nestor KE, Saif YM: **Genetic association of selection for increased leg muscle and increased shank diameter with body composition and walking ability.** Poultry Sci 1991, **70:** 739-745.

40. Ankra-Badu GA, Aggrey SE: **Identification of candidate genes at quantitative trait loci on chicken chromosome Z using orthologous comparison of chicken, mouse and human genomes.** In Silico Biol. 2005, **5:** 593-604.

41. Gulberg D, Tiger CF, Velling T: **Laminins during muscle development and in muscular dystrophies.** Cell Mol Life Sci 1999, **56:** 442-460.

42. Gawlik K, Miyagoe-Suzuki Y, Ekblom P, Takeda S, Durbeej M: **Laminin α1 chain reduces muscular dystrophy in laminin α2 chain deficient mice.** Hum Mol Genet 2004, **13:** 1775-1784.

43. Kolsch H, Lutjohann D, Tulke A, Bjorkhem I, Rao ML: **The neurotoxic effect of 24-hydroxycholesterol on SH-SY5Y human neuroblastoma cells.** Brain Res 1999, **818:** 171-175.

44. Hayden JM, Brachova L, Higgins K, Obermiller L, Sevanian A, Khandrika S, Reaven PD: **Induction of monocyte differentiation and foam cell formation in vitro by 7-ketocholesterol.** J Lipid Res 2002, **43:** 26-35.

45. Crosbie RH, Dovico SA, Flanagan JD, Chamberlain JS, Ownby CL, Campbell KP: **Characterization of aquaporin-4 in muscle and muscular dystrophy.** FASEB J 2000, **16:** 943-949.

46. Li MD, Kane JK: **Effect of nicotinic on the expression of leptin and forebrain leptin receptors in the rat.** Brain Res 2003, **991:** 222-231.

47. Schmid M, Nanda I, Guttenbach M, Steinlein C, Hoehn M, Schartl M. et al: **First report on chicken genes and chromosomes 2000**. Cytogenet Genome Res 2000, **90:** 169-218.

48. Iannucelli E, Woloszyn N, Arhainx J, Gellin J, Milan D: GEMMA: A database to automate microsatellite genotyping, Proceedings of the XXVth International Conference on Animal Genetics, Tours-France, 21-25 July 1996. Anim Genet 1996,27, **Suppl. 2:**55

49. Haley CS, Knott SA, Elsen JM: **Mapping quantitative trait loci in crosses between outbred lines using least squares.** Genetics 1994, **136:** 1195-1207.

50. Seaton G, Haley CS, Knott SA, Kearsey M, Visscher PM: **QTL Express: mapping quantitative trait loci in simple and complex pedigrees.** Bioinformatics 2002, **18:** 339-340.

51. Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping.** Genetics 1994, **138:** 963-971.

52. Lander E. and Kruglyak L**: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.** Nat Genet 1995, **11:** 241-246.

Table 3.1 Means and standard deviations of body weight, growth and skeletal traits of F_2 individuals from reciprocal F_1 crosses of a divergent chicken line selected for fast (HG) or slow growth (LG).

	Mean ± SD		
Trait	(LH ♂line) (N=280)	(HL ♂line) (N=415)	(Combined) N=695
BW0, g	36.56± 2.19	34.90± 2.32	35.57± 2.36
BW1, g	63.07± 10.63	59.04± 9.81	60.66± 10.33
BW3, g	230.92± 29.26	229.45± 33.00	230.04± 31.53
BW5, g	472.38± 70.90	487.60± 75.65	481.47± 74.10
BW7, g	743.09±112.07	783.32±125.64	767.11±121.89
BW9, g	1087.40±177.05	1154.70±186.69	1127.59±185.70
SL, mm	98.27± 6.78	97.96± 6.58	98.09± 6.66
SD, mm	8.91± 0.84	9.10± 0.82	9.02± 0.83
Growth1, g	26.65± 9.70	24.14± 9.73	25.15± 9.80
Growth2, g	166.90± 26.02	170.41± 26.69	169.00± 26.46
Growth3, g	241.46± 48.22	258.15± 51.27	251.43± 50.69
Growth4, g	270.71± 60.74	296.74± 64.82	286.25± 64.45
Growth5, g	344.31± 79.14	368.72± 91.74	358.89± 87.65

BW0=body weight (BW) at hatch; BW1=BW at week 1; BW3=BW at week 3; BW5=BW at week 5; BW7=BW at week 7; BW9=BW at week 9; SL=shank length; SD=shank diameter; *Growth1*=BW1-BW0; *Growth2*=BW3-BW1; *Growth3*=BW5-BW3; *Growth4*=BW7-BW5; *Growth5*=BW9-BW7.
LH (LG ♂x HG ♀); HL (HG ♂x LG ♀)

Table 3.2 Number of markers, map length and first marker for each chromosome (linkage group)

Chromosome	Number of markers used	Map length (cM)	First Marker	Last Marker
1	18	558	MCW0168	MCW0108
2	17	432	MCW0205	MCW0157
3	11	289	ADL0177	MCW0037
4	6	233	ADL0317	LEI0037
5	8	141	LEI0082	ADL0298
6	4	93	LEI0192	ADL0323
7	6	151	ADL0315	LEI0064
8	3	54	LEI0136	MCW0305
9	3	71	LEI0028	ADL0132
10	4	53	LEI0112	MCW0067
11	3	49	MCW0097	ADL0308
12	3	31	ADL0372	ADL0044
13	3	35	MCW0110	MCW0213
14	2	43	MCW0123	MCW0123
15	3	50	MCW0211	ADL0206
17	1	21	ADL0293	ADL0199
18	2	17	ADL0304	MCW0217
19	2	15	MCW0266	MCW0256
26	3	29	ADL0330	LEI0074
27	3	13	MCW0146	MCW0233

Table 3.3 Growth traits, QTL positions, markers and confidence intervals for each chromosome.

			Map Position			
GGA	Trait	Flanking Markers	(cM)	Sequence Map (bp)	95% C.I.	F-Ratio
1	BW7	LEI0209-LEI0252	73	16598396-36369954	15-558	3.99^{+}
1	*Growth4*	LEI0209-LEI0252	78	16598396-36369954	16-558	4.82^{+}
1	SL	ADL0328-LEI0061	482	161290914-174817966	80-501	4.95^{+}
2	BW5	LEI0147-MCW0096	301	97299246-104970323	72-381	6.29*
2	*Growth3*	LEI0147-MCW0096	302	97299246-104970323	17-390	5.10^{+}
2	*Growth2*	MCW0096-MCW0264	314	104968302-112313816	8-381	4.49^{+}
2	SL	MCW0096-MCW0264	316	104968158-112309159	105-391	4.57^{+}
2	BW9	MCW0056-LEI041	375	124151152-133390996	10-391	5.03^{+}
2	BW7	MCW0056-LEI041	379	124151152-133390996	11-402	5.21^{+}
2	*Growth4*	MCW0056-LEI041	382	124151152-133390996	4-419	4.68^{+}
3	BW0	ADL0371	87	36388876-36392890	0-279	4.33^{+}
3	*Growth4*	ADL0280-ADL0306	170	54765967-80340021	27-278	6.04**
3	BW7	ADL0280-ADL0306	176	54765967-80340021	19-279	4.82^{+}
3	BW9	ADL0280-ADL0306	179	54765967-80340021	7-284	3.57^{+}
3	SD	ADL0280-ADL0306	184	54765967-80340021	0-276	4.70^{+}
4	*Growth5*	MCW0240-LEI0073	200	68517805-80277860	186-214	14.35**
4	SL	MCW0240-LEI0073	207	68517805-80277860	198-214	34.11**
4	*Growth3*	MCW0240-LEI0073	208	68517805-80277860	193-217	15.39**
4	BW9	MCW0240-LEI0073	208	68517805-80277860	195-216	19.20**
4	SD	MCW0240-LEI0073	209	68517805-80277860	202-215	40.07**
4	BW5	MCW0240-LEI0073	212	68517805-80277860	185-223	11.15**
4	BW7	MCW0240-LEI0073	212	68517805-80277860	197-221	16.58**
4	*Growth4*	MCW0240-LEI0073	212	68517805-80277860	200-222	16.51**
4	BW3	MCW0240-LEI0073	218	68517805-80277860	4-233	3.62^{+}
4	*Growth2*	MCW0240-LEI0073	218	68517805-80277860	0.5-233	3.73^{+}
5	BW7	MCW0193-MCW0214	37	12359652-24784335	9-136	4.30^{+}
5	BW9	MCW0193-MCW0214	39	12359652-24784335	9-126	4.43^{+}
5	*Growth4*	MCW0193-MCW0214	39	12359652-24784335	9-103	5.61*
9	*Growth5*	LEI0028-LEI0130	24	8181815-13729032	0-71	3.33^{+}
11	*Growth2*	ADL0210-ADL0308	42	12225839-13444395	0-49	2.67^{+}
12	*Growth4*	ADL0372	0	668688-670706	0-31	3.42^{+}
12	*Growth5*	ADL0372-LEI0099	11	668688-2299449	0-31	3.18^{+}
13	*Growth5*	ADL0372-LEI0099	15	982361-8512843	0-35	3.96^{+}
13	SD	MCW0197-MCW0213	20	982361-8512843	2-35	4.69^{+}
13	BW9	MCW0197-MCW0213	22	982361-8512843	0-35	4.14^{+}
13	*Growth4*	MCW0213	35	981341	0-35	3.40^{+}
17	*Growth1*	ADL0293	0	5049779-5051798	0-21	3.27^{+}
17	*Growth3*	ADL0199	21	10477404-10481423	0-21	2.33^{+}
17	*Growth5*	ADL0199	21	10477404-10481423	0-21	2.77^{+}
18	*Growth5*	ADL0304-MCW0217	12	1404975-3045524	0-17	3.19^{+}
26	BW7	ADL0330-MCW0069	6	41662-1208634	0-25	4.09^{+}
26	BW9	ADL0330-MCW0069	7	41662-1208634	0-20	4.36^{+}
26	*Growth4*	ADL0330-MCW0069	7	41662-1208634	0-22	4.00^{+}
26	*Growth3*	ADL0330-MCW0069	8	41662-1208634	0-27	2.96^{+}

BW0=body weight (BW) at hatch; BW1=BW at week 1; BW3=BW at week 3; BW5=BW at week 5; BW7=BW at week 7; BW9=BW at week 9; *Growth1*=BW1-BW0; *Growth2*=BW3-BW1; *Growth3*=BW5-BW3; *Growth4*=BW7-BW5; *Growth5*=BW9-BW7; SL=Shank length; SD=Shank diameter. +Suggestive linkage; *Significant linkage (P<0.05) genome wide; **Significant linkage (P<0.01) genome wide

Table 3.4 Additive and Dominance effects and the percent variance (PV) explained by the QTL.

GGA	Trait	Male		Female		PV[2]
		Additive Effects	Dominance Effects	Additive Effects	Dominance Effects	
1	BW7	28.72 (10.43)	32.22 (25.22)	29.74 (11.82)	-21.72 (26.64)	2.0
1	*Growth4*	15.09 (5.43)	29.87 (13.05)	13.89 (6.11)	-14.30 (13.75)	3.0
1	SL	1.00 (0.36)	0.79 (0.53)	0.68 (0.38)	-1.28 (0.57)	3.0
2	BW5	16.90 (5.49)	-0.06 (11.50)	15.60 (5.93)	32.51 (12.41)	4.0
2	SL	1.10 (0.47)	-0.41 (1.02)	1.41 (0.50)	2.20 (1.09)	3.0
2	*Growth3*	10.36 (3.47)	1.05 (7.20)	7.56 (3.75)	19.04 (7.77)	3.0
2	*Growth2*	6.00 (2.31)	-1.68 (4.96)	6.64 (2.47)	9.76 (5.31)	3.0
2	BW9	46.05 (14.05)	36.55 (27.46)	36.49 (14.21)	33.13 (29.73)	3.0
2	BW7	25.70 (8.91)	34.98 (16.77)	25.14 (9.06)	8.27 (17.21)	3.0
2	*Growth4*	11.78 (4.41)	16.49 (7.89)	11.66 (4.52)	6.28 (8.31)	3.0
3	BW0	0.26 (0.13)	-0.50 (0.22)	-0.03 (0.15)	0.71 (0.25)	3.0
3	Growth4	31.05 (7.79)	-37.45 (25.82)	0.33 (9.24)	-72.94 (28.74)	4.0
3	BW7	48.99 (14.59)	-77.70 (48.42)	4.85 (17.26)	-128.05 (54.85)	3.0
3	BW9	58.88 (21.81)	-108.27 (71.33)	19.50 (25.69)	-160.12 (81.26)	2.0
3	SD	0.32 (0.09)	-0.43 (0.28)	-0.81 (0.11)	-0.50 (0.32)	3.0
4	*Growth5*	29.02 (5.43)	-4.61 (11.42)	33.69 (6.38)	-5.05 (12.16)	8.0
4	SL	3.72 (0.41)	-0.34 (0.93)	3.37 (0.47)	-1.29 (1.00)	18.0
4	*Growth3*	18.08 (3.19)	1.61 (7.30)	19.64 (3.70)	-6.56 (7.88)	8.0
4	BW9	82.71 (13.04)	19.22 (29.82)	89.56 (15.14)	-20.70 (32.22)	11.0
4	SD	0.53 (0.05)	-0.12 (0.13)	0.50 (0.06)	-0.23 (0.14)	21.0
4	BW5	23.61 (52.23)	14.76 (12.10)	27.14 (6.01)	-15.81 (13.61)	7.0
4	BW7	51.12 (8.86)	24.04 (20.48)	55.97 (10.18)	-16.81 (22.29)	9.0
4	*Growth4*	27.51 (4.62)	9.27 (10.68)	28.84 (5.31)	-1.00 (11.63)	9.0
4	BW3	4.82 (2.71)	11.31 (6.11)	7.86 (3.06)	-6.45 (6.70)	2.0
4	*Growth2*	3.80 (2.31)	9.31 (5.21)	7.31 (2.61)	-5.27 (5.71)	2.0
5	BW7	22.77 (8.31)	-11.02 (16.33)	29.75 (9.66)	15.47 (17.78)	3.0
5	BW9	29.46 (12.43)	-8.43 (24.00)	50.44 (14.48)	25.98 (26.14)	3.0
5	*Growth4*	14.23 (4.23)	-1.47 (8.17)	15.55 (4.93)	14.79 (8.90)	3.0
9	*Growth5*	0.07 (5.40)	-27.24 (9.24)	12.76 (6.0)	-0.73 (9.82)	2.0
11	*Growth2*	3.60 (1.97)	-5.97 (3.50)	-0.69 (2.12)	-8.32 (3.78)	2.0
12	*Growth4*	0.51 (3.47)	15.08 (4.93)	-7.53 (3.62)	-0.29 (5.32)	2.0
12	*Growth5*	-5.70 (5.98)	26.41 (10.63)	-11.77 (6.24)	-17.09 (11.71)	2.0
13	*Growth5*	19.18 (4.99)	-5.75 (7.79)	5.28 (5.21)	-0.51 (8.04)	2.0
13	SD	0.19 (0.05)	0.03 (0.80)	0.10 (0.05)	0.09 (0.08)	3.0
13	BW9	43.80 (11.46)	-17.29 (18.48)	16.11 (12.01)	0.91 (19.09)	3.0
13	*Growth4*	12.54 (3.73)	-3.74 (5.52)	5.91 (4.02)	1.54 (6.00)	2.0
17	*Growth1*	-1.92 (0.66)	2.94 (1.48)	0.32 (0.72)	1.11 (1.58)	2.0
17	*Growth3*	7.05 (3.35)	8.08 (6.90)	3.46 (3.71)	-11.68 (7.57)	1.0
17	*Growth5*	13.57 (6.26)	-23.01 (12.88)	4.60 (6.92)	-25.17 (14.14)	2.0
18	*Growth5*	14.10 (5.93)	-25.41 (10.31)	5.92 (6.69)	-2.33 (11.33)	2.0
26	BW7	9.62 (7.29)	-24.46 (11.89)	26.78 (8.40)	-0.07 (12.59)	3.0
26	BW9	19.49 (11.10)	-25.79 (18.01)	44.54 (12.78)	-5.44 (19.19)	3.0
26	*Growth4*	5.04 (3.80)	-12.36 (6.16)	13.78 (4.37)	-0.28 (6.57)	2.0
26	*Growth3*	3.10 (2.68)	-6.09 (4.30)	8.54 (3.09)	2.36 (4.63)	2.0

[2] % Phenotypic variance explained by the QTL; [3] standard errors in parenthesis

BW0=body weight (BW) at hatch; BW1=BW at week 1; BW3=BW at week 3; BW5=BW at week 5; BW7=BW at week 7; BW9=BW at week 9; *Growth1*=BW1-BW0; *Growth2*=BW3-BW1; *Growth3*=BW5-BW3; *Growth4*=BW7-BW5; *Growth5*=BW9-BW7; SL=Shank length; SD=Shank diameter

Table 3.5 QTL by sex interaction for carcass traits in an F_2 population derived from a divergent chicken line selected for high or low growth

Trait[1]	F-Ratio	Location (cM)[2]	Additive Effect	Dominance effect	PV[3]
GGA3					
BW7					
Male	6.43*	173 (75-228)	42.05 ±13.21	-57.98 ± 36.45	3.28
Female	3.06	172 (7-287)	-21.23 ± 8.68	-4.53±15.20	1.49
Growth4					
Male	7.42*	171 (188-288)	25.96 ± 7.32	-30.08 ±20.46	3.92
Female	2.96	174 (21-287)	-9.98 ± 4.32	-5.47 ± 7.58	1.41

[1]BW7=Body weight (BW) at 7weeks; *Growth*4=BW7-BW5;
*Significant at genome-wide level at P≤0.05
[2]95% confidence interval in brackets
[3]Proportion of the phenotypic variance explained by QTL

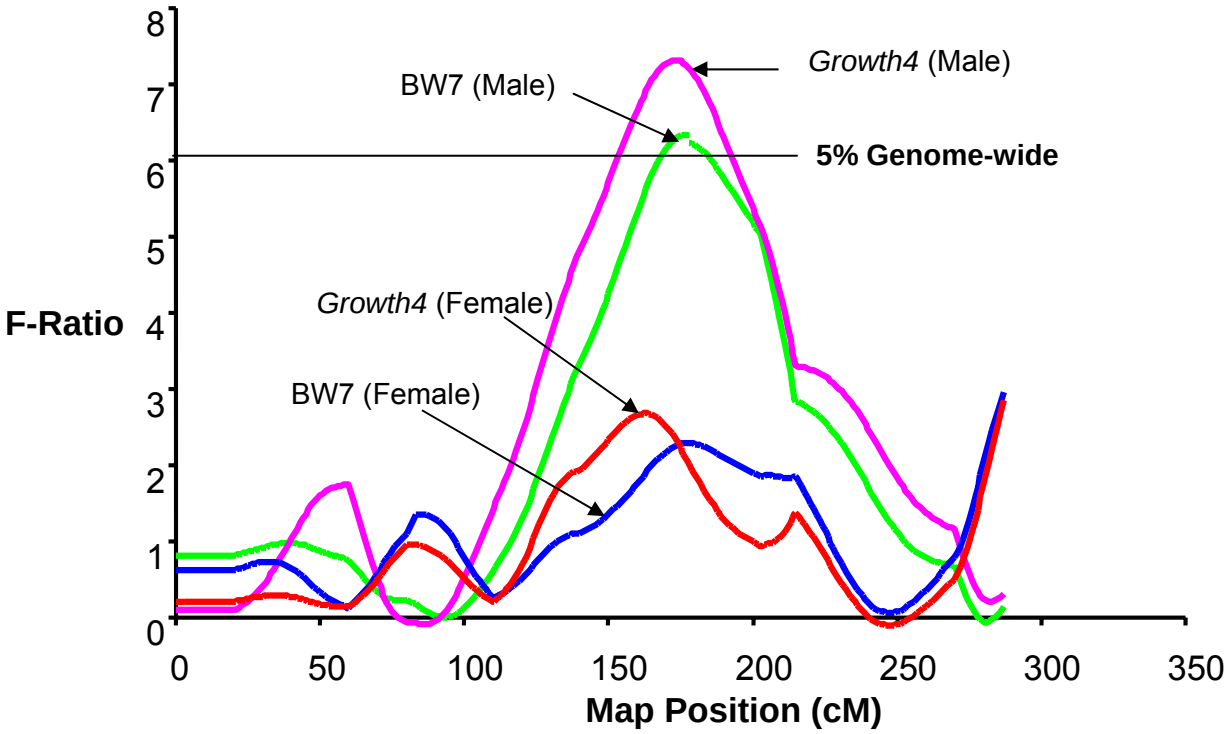

Figure 1.Sex-specific Body Weight and Growth QTL on GGA3

CHAPTER 5

MOLECULAR DISSECTION OF FATNESS AND BODY COMPOSITION IN A DIVERGENT CHICKEN LINE SELECTED FOR HIGH AND LOW BODY WEIGHT[1]

[1] Georgina A. Ankra-Badu and Samuel E. Aggrey. To be submitted to Animal Genetics

ABSTRACT

Genes influencing carcass composition and fatness were mapped in an intercross of a divergent line selected for fast or slow growth. Genome-wide significant loci effects for *Pectoralis* (*P*) *major* muscle were found on GGA 2, 3, 4, and 10, for *P. minor* muscle on chromosome 1, 2, and 4, for thigh weight on GGA2, 4, 5, and 26, and for abdominal fatness on chromosomes 1, 4, and 5. Evidence of female specific quantitative trait loci (QTL) were found for fatness and *P. major* percentage GGA1, and for thigh weight on GGA26. A male specific QTL for *P. major* weight was detected on GGA3. The strongest QTL was found on GGA4 that accounted for about 0.5 standard deviations respectively in breast and thigh weights and abdominal fatness. Differentially expressed genes within the QTL location on GGA4 included superoxide dismutase (*SOD*) and FAT tumor suppressor (*Drosophila*) homolog (*fat-1*). Some of the genes identified near the GGA4 QTL are involved in embryogenesis and muscle development, although no associations have been reported with carcass traits. Thyroid hormone responsive Spot 14 (*THRSP*), a transcription factor, is located near the QTL for fatness on GGA1. Duplicated polymorphic paralogs of Spot 14, *THRSPα* and *THRSPβ* have been found to be associated with abdominal fat in chickens.

INTRODUCTION

A principal goal of genetics is the determination of quantitative genetic variation in traits and the complex genetic architecture of the underlying factors that explains the variation in traits. The genetic basis of many phenotypes takes the form of a continuous range rather than a discrete class. The complexity of phenotypes exhibiting continuous variation

often results from the segregation of many genes, with minor and major effects, and whose expression are modified by both the environment, artificial and/or natural selection, and genetic background. The discovery of genes and the polymorphisms could explain the quantitative variation in complex traits. This will require information obtained from various methods to unlock the full potential of genomes available. Quantitative trait loci (QTL) mapping has been widely used to determine loci that are responsible for variation in complex quantitative traits.

In chickens, crosses between extreme strains and/or breeds have been used to detect QTL that explain quantitative variation between the parental populations. Several studies have reported QTL for body weight and feed efficiency (van Kaam et al. 1998; Sewalem et al. 2002; de Koning et al. 2004), egg production and quality (Tuiskula-Haavisto et al. 2002; Sasaki et al. 2004), behavior (Schűtz et al. 2002) and fatness (Ikeobi et al. 2002; Jennen et al. 2004; Aggrey et al. 2005). Identification of genes controlling these traits in chickens has a dual purpose; an obvious application for poultry improvement, and for biomedical research as a model organism. Lean and fatness (obesity) are at opposite ends of a continuous distribution of fatness (Pomp 1997). The genetic contribution to obesity as estimated by heritability of fatness phenotypes, varies from 30 to 70% (Comuzzie and Allison 1998). Identification of specific genes or QTL that are linked to fatness is of interest in poultry breeding and human health.

The dissection of QTL at the nucleotide level, quantification of epistasis, and evaluation of genotype by environmental interactions are difficult to identify (Gibson and Mackay 2002). The confidence interval (CI) for a QTL position is usually large and can harbor thousands of genes. Fine mapping can reduce the number of candidate genes to a

few hundred. Even after fine mapping, there could be more candidate genes than one can tentatively pursue. Moreover, almost all fine mapping approaches have their own drawbacks (Darvasi 1998). There is a need to further reduce candidate genes in QTL regions that have a high probability of becoming actual genes involved in the architecture of the traits. Quantitative expression of genes within QTL regions is one method that can be used to reduce the number of putative candidate genes for further study.

In an attempt to identify candidate genes involved in the genetic architecture of fatness and meat quality in the chicken, we have combined conventional QTL mapping with gene expression profiling. This integrative approach has the potential to identify genes causally involved in expressing of complex traits.

RESULTS

QTL Effects

The QTL affecting carcass traits in our F_2 resource population are summarized in Table 2. Twenty-seven QTL with suggestive and significant linkages were observed on 8 chromosomes. The QTL positions, microsatellite markers flanking the QTL and the F ratios are also shown in Table 2. Highly significant QTL for breast meat weight, thigh weight, *P. major* and *P. minor* weights were found on GGA4 around the same location. A significant abdominal fat weight QTL was also found (at the chromosome level) at the same location on GGA4. Abdominal fat QTL were found on GGA1 and GGA5. Additional significant QTL for thigh weight were identified on GGA2, GGA5 and GGA26. *P. minor* weight QTL was also found GGA2. There were suggestive QTL for breast meat weight on GGA1-3, for *P. major* weight on GGA2-3, and for *P. minor*

weight on GGA1. Adjusting carcass traits for BW at 9 wk revealed additional significant QTL on GGA1, 3 and 6 for breast meat percentage, and also abdominal fatness on GGA4. The additive effects of the significant QTL with effect of the phenotypic standard deviation explained by the QTL and the phenotypic variance explained by the QTL for both males and females are shown in Table 3. Notably the QTL on GGA4 (located between *MCW240* and *LEI073*) explains about 0.5 of the phenotypic standard deviation for the traits affected in both sexes. The QTL effects ranged from 2.02 to 10.40% of the phenotypic variance.

Interaction of QTL with sex was significant for some traits. The additive effect of QTL by sex interaction is shown in Table 4. The QTL effect on GGA1 for breast meat percentage, fat weight and fat percentage, and *P. major* percentage were significant at the genome-wide level in the females but not in males. A similar observation was made on GGA26 for thigh weight. However, on GGA3, the QTL for *P. major* weight was significant in males, but not females. After fitting two QTL for each trait on a chromosome, there was evidence of two QTL affecting *P. major* yield on GGA1 (221 cM and 485 cM). . There was no evidence of dominance or imprinting in any of the traits analyzed.

DISCUSSION

QTL for Carcass Traits

This study reveals a number of significant QTL for carcass traits in an F_2 resource population established from commercial meat-type birds and divergently selected for either fast (FGL) or slow growth (SGL). Carcass traits were adjusted for using body

weight at 9 weeks of age. The results were similar when BW at 9 weeks was used as a covariate for carcass measurements. Chicken breast muscle is comprised of two muscle groups: *P major* and the *P. minor*. Significant QTL for total breast weight were detected on GGA4 and breast meat percentage on GGA1, GGA3 and GGA6. The *P. major* was the contributing factor for the breast muscle QTL on GGA1 at (221 cM), however, GGA1 contains an additional QTL for *P. major* muscle (at 485 cM). The *P. major* weight QTL confirms the breast meat weight QTL reported by Lagarrigue et al. (2006) in the same region. Our QTL analysis shows that both *P. major* muscle QTL on GGA1 are female specific. A suggestive QTL for breast meat weight on GGA3 was identified (at 176 cM), however, sex by trait analysis revealed that the significant QTL for breast meat weight was male specific and primarily due to *P. major* weight (at 172 cM). We found a novel QTL for breast meat on GGA4 that accounted for 0.5 standard deviation of the phenotypic variation in both sexes. This major QTL affects both *P. major* and *P. minor* muscle weight. This implies that identification of a candidate gene responsible for the QTL on GGA4 has the potential to simultaneously improve the *P. major* and *P. minor* weights in both sexes. Another significant QTL affecting *P. minor* exclusively was observed on chromosome 2. McElory et al. (2002) reported a total breast muscle QTL on GGA2, but at a different location. Chicken breast muscle represents the most important carcass trait in broiler chickens because of the premium paid by consumers. The *P. major* is usually sold as filet, while the *P. minor* as the 'chicken finger'. Heritability estimates for breast meat have been reported as a single trait (Le Bihan-Duval et al. 2001; Zerehdaran et al. 2004). The current study suggests similar genes could affect both *P. major* and *P. minor* muscles. However, other genes could control them independently

therefore, combining them as one trait could introduce a bias in the estimation of genetic parameters. Four QTL that affect thigh weight are located on GGA2, 4, 5 and 26 that explain about 17% of the phenotypic variation attributed to the FGL allele. The heritability of thigh weight ranges 0.31 to 0.7 (Ricard and Rouvier 1969; Cahaner and Nitsan, 1985) with positive genetic correlations with other carcass measurements. The QTL for thigh weight is different from the one reported by Ikeobi et al. (2004) from a broiler-layer cross at position 0 cM on GGA4. The QTL for thigh weight on GGA4 alone explains about one-third of the phenotypic standard deviation of that trait. Identification the underlying gene(s) and development of polymorphic markers in those genes will be the first step towards marker-assisted selection (MAS). The additive QTL effect of thigh weight on GGA4 is about 10 g. It should be noted that this value should be doubled (i.e., 20 g) to estimate the effect of replacing the SGL allele with the FGL allele, which represents a QTL of significant economic importance. During the last decade whole poultry carcasses are mainly sold as processed products. Therefore, an understanding of the molecular basis of meat yields would enable poultry breeders to select more appropriate breeding stocks for further meat yields.

QTL for Fatness Traits

Genetic selection over the last 50 years in meat-type chicken has led to rapid growth and a concomitant increase accretion of body fat (Havenstein et al. 2003). Fatness is a complex trait affected by both genes and the environment (i.e., nutrition, appetite, behavior, etc.), and their interactions. The genetic basis of fatness is of great interest to both poultry production and human health. Our study shows a significant QTL for abdominal fat was observed on GGA1 and GGA5 while a suggestive QTL for abdominal

fat was observed on GGA4. However, a highly significant QTL for fatness was identified after adjustment for BW at 9wk of age. The fatness QTL on GGA4 is attributed to the SGL allele since it causes a reduction in the percentage abdominal fat. The QTL for fatness on GGA1 is located within the confidence interval of a QTL for skin fat reported by Ikeobi et al. (2002) and Jennnen et al. (2004). Our analysis shows that the QTL for fatness on GGA1 is female specific. The QTL for fatness on GGA5 confirms a similar QTL reported earlier (Ikeobi et al. 2002; Lagarrigue et al. 2006). Sex-specific QTL have been mapped for longevity in *Drosophila melanogaster* (Nuzhdin et al. 1997), skeletal muscle weight (Lionikas et al. 2003) and fatness (Jerez-Timaure et al. 2004) in mice, and fatness in chicken (Abasht et al. 2006). This display of sexual dimorphism is common in most vertebrates, since the male and female follow different growth trajectories. The underlying factors that govern sex-influenced traits could be different, and this could be a reflection of the genetic basis for the same trait in the sexes (Aggrey and Cheng 1994). The results from the present study indicate that it is important to differentiate between the sexes in a QTL analysis, since three female specific QTL (breast meat percentage, fat yield, and thigh weight) were found in female chickens, and one male specific QTL for thigh weight was found in male chickens. To identify sex-specific QTL, analysis of the entire population with and without sex-by-QTL interaction should be conducted. Failure to detect a significant interaction suggests a common QTL. A separate male and female analysis should only be performed when there is substantial evidence to support the claim. The mechanism underlying sex-specific effects are unknown but may arise from the influence of sex hormones on the regulation of the genes that underlie these QTL.

These studies highlight the importance of in taking sexual dimorphism into account in the analysis of fatness QTL in both animal models and human studies.

The genetic effect of multiple QTL identified on GGA4 is complex. The FGL QTL allele has additive increasing thigh and breast meat weights, while the SGL QTL allele has a decreasing additive effect on fatness. Breast meat weight and fatness have moderate to high heritabilies and negative genetic correlation between them (Zerehdaran et al. 2004). It is plausible that the region (200-220 cM) on GGA4 harbors at least 2 QTL that are linked together and acting in opposite direction. Such a relationship would be favorable for breeding meat-type birds since a reduction in fatness and increase in breast and thigh mass is most desirable.

The recurrence of similar QTL in different crosses suggests a limited number of key genes involved in the phenotype. However, some QTL may result from the evolutionary pressure, artificial or natural on the parental strains and their adaptive environment(s). Consequently, the genetic architecture of such QTL for the same phenotype may be different.

From QTL to gene:

A major goal in fine QTL mapping is the identification of the genes that underlie polygenic traits of importance. The QTL identified on GGA4 has the strongest effect with the smallest confidence interval. Some putative candidate genes that could be associated with QTL traits are presented in Table 5. Many of these genes are involved in some aspects of muscle development. Differentially expressed genes within the QTL region could be candidates that underlie QTL. Prioritizing specific candidate genes from transcriptional profiles is very beneficial when the location of the QTL is known with

high degree of certainty. The novel QTL region on GGA4 shows conservation of synteny with parts of HSA4p15.1 and MMU 5B3, 5C1 and 5C2 (http://www.ensembl.org/Gallus gallus). Analysis of the othologous region (GGA4) with human and mouse chromosomes provides some additional candidate genes that have not been mapped onto the chicken genome. This approach was used by Ankra-Badu and Aggrey (2005) to identify novel candidate genes for QTL located on chicken chromosome Z.

Potential candidate genes mapped in the QTL region on GGA4 are fibroblast growth factor 2 (*FGF 2*), bone morphogenetic protein receptor type IB (*BMPR1B*), alpha actinin associated LIM protein smooth muscle isoform (*PDZ and LIM*), fibroblast growth factor receptor 3 (*FGFR3*), fibroblast growth factor receptor-like 1 (*FGFRL1*), platelet-derived growth factor receptor alpha (*PDGR-A*), *SLIT2*, v-kit Hardy-Zukerman 4 feline sarcoma viral oncogene homolog (*KIT*), tec protein tyrosine kinase (*TEC*), actin filament associate protein (*AFAP*), neuronal acetylcholine receptor subunit alpha 4, extracelluar superoxidase dismutase (*EC-SOD*), alcohol dehydrogenase-1 (*ADH-F*), and FAT tumor suppressor homolog 1 (*FAT-1*). Alpha actinin is required for the organization and function of the contractile machinery of muscle. Alpha-actinin-binding PDZ-LIM protein is dramatically upregulated during smooth and skeletal muscle differentiation and are associated with α-actinin at key sites for muscle cyto-architecture (Pomies et al. 1999). The basic fibroblast growth factor (b*FGF 2*) belongs to a multigene family consisting of seven related members (FGF-1 to –7) that are evolutionary very conserved. A common feature of FGFs is their affinity to heparin, which reflects their binding to the extracellular matrix. The FGFs are involved in cell differentiation, regulation of cell proliferation, and progression through cell cycle (Borja et al. 1992). Other fibroblast

growth factor related genes *FGFR3, FGFBP, and FGFRL-1* around the same QTL location play critical roles in embryogenesis. During later embryogenesis bFGF is involved in muscle morphogenesis (Kardami and Fandrich, 1989).

Other potential candidates within the QTL regions on GGA1 that affect fatness and *P. major* muscle (435 cM and 539 cM) are frizzled homolog 4 (Drosophila) (*FZD4*), thyroid hormone responsive spot 14 (*THRSP*), uncoupling protein 3 (*UCP3*), parathyroid hormone-like hormone (*PTHLH*), pyrimidinergenic receptor (*P2RY6*), integrin-linked kinase (*ILK*), signal peptidease complex subunit 2 homolog, cyclin D2 (*CCND2*), Rho guanine nucleotide exchange factor (*ARHGFF17*), mitochondrial ribosomal protein L48 (*MRPL48*), G protein-coupled P2 receptor (*P2Y3*).

The *THRSP* gene encodes a small acidic protein implicated as transcription factor involved in lipogenic enzymes and highly expressed in lipogenic tissues (i.e., liver, fat and the mammary gland) (Jump and Oppenheimer 1985). The *THRSP* gene is differentially expressed in the FGL and SGL and it has been shown that the mRNA expression is regulated by the thyroid hormone status in meat-type chickens (Wang et al. 2002). Polymorphic paralogs of Spot 14 (*THRSPα* and *THRSPβ*) are associated with abdominal fat in chickens (Wang et al. 2004).

As there are many genes within each of the QTL regions, most of unknown functions, resolving the actual loci contributing to variation of the traits studied will require additional mapping efforts. Identification of the causative genes to the traits studied will require fine-scaled mapping of the QTL regions and complementation tests to candidate genes within refined QTL regions (Pasyukova et al. 2000). Any inferences from putative candidate genes should be done with caution since the confidence interval

associated with the QTL are generally large. Strategies that can be used to further reduce the confidence interval of QTL locations have been described by Darvasi (1998). Nevertheless the combination of transcriptional profiling, comparative mapping using the sequence of the chicken genome and QTL analysis increase the chance to identify potential candidate genes whose allelic variants may be involved in the genetic architecture of polygenic traits of importance.

In summary the QTL identified in this study have strong additive effect on breast and thigh weight, and abdominal fatness. Since the divergent lines used in this study were derived from commercial lines, it is still possible that such QTL are segregating in today's commercial stocks. Identification of genes underlying these QTL could aid selection. Human medicine could also benefit from deciphering fatness genes as obesity has a strong association with hypertension, diabetes and other cardiac related disease.

MATERIALS AND METHODS

Animals and Phenotypes

An F_2 population was generated by inter-mating two experimental boiler lines that were divergently selected for fast (FGL) or slow growth (SGL) (Ricard 1975). In the F_0 generation, five males from each line were mated with 2-5 females of the alternative line to generate the F_1 population. The F_1 males were reciprocally mated with females to generate 695 F_2's. The F_2 population were produced in four hatches, fed on standard broiler diets *ad libitum* (0-3 wk; 3050 kcal ME, 4-9 wk; 3100 kcal ME), and raised under standard management practices for nine weeks. Blood was taken from all birds for DNA extraction. At 9-wk, birds were weighed after an overnight fast and slaughtered. After

evisceration, carcasses were stored overnight at 4^{o}C before dissection. The carcass traits measured are total breast meat, *Pectoralis* (*P) major*, *P. minor*, abdominal fat and thigh weights.

Genetic markers and Genotyping

DNA was extracted from 100 µl of whole blood using phenol-chloroform extraction. Microsatellite markers were selected from the poultry genetic consensus map (Schmid et al. 2000) based on their location on the chromosome and informativeness in each F_0 sire family. One hundred and seven markers from 20 autosomal linkage groups and the sex chromosomes were used for genotyping (Table 1). Some markers were combined (2~10) according to their size and amplification conditions for multiplex PCR amplification. Fluorescently labelled microsatellite sequences were analysed on ABI 3700 DNA sequencer (Applied Biosystems, Foster City, CA, USA) and their lengths were determined using GeneScan Analysis™ software (Applied Biosystems, Foster City, CA, USA). Genotypes were interpreted using both the Genotyper Analysis™ software (Applied Biosystems, Foster City, CA, USA) and GEMMA databases (Iannuccelli et al. 1996).

Statistical Analysis and QTL mapping

Breast meat, *P. major*, *P. minor*, thigh and abdominal fat weights were adjusted with BW at wk 9 to obtain their respective percentages for further analysis. The linkage map of the 20 autosomal linkage groups and the sex chromosome were generated using the CRI-MAP linkage program with marker distances in Kosambi centiMorgans (cM) (Green et al. 1990). The linkage map based on 107 marker loci was in good agreement with the chicken consensus map (Schmid et al. 2000). The QTL interval mapping method of

Haley et al. (1994) was implemented using QTL Express software (Seaton et al. 2001), using the F_2 model option. Marker genotypes were used to estimate the probabilities of line origins of each gamete at 1 cM intervals throughout the genome for each F_2 individual. The conditional probability of an F_2 progeny being each of the four QTL genotypes (QQ, Qq, qQ, and qq) was calculated. The least square regression model used for QTL analysis included the fixed effects of sex, hatch and family ($Y_{ijkl} = \mu + hatch_i + Sex_j + QTL_k + (QTL \times Sex)_{jk} + e_{ijkl}$). A model for the additive (*a*) and dominance (*d*) effect of a QTL, and QTL by sex interaction at a given position was used for the analysis; the model was run separately for males and females. A model fitting an imprinting (paternal origin of allele) effect was tested as described by Knott et al. (1998). Statistical significance threshold for QTL presence was determined by 1000 empirical permutations of the data (Churchill and Doerge 1994). Genome-wide threshold ($\alpha = 0.05$ and $\alpha = 0.01$) for significant QTL and chromosome-wide threshold of $\alpha = 0.05$ for suggestive QTL were applied. Confidence intervals (CI) for QTL location were calculated from 1000 boostrap samples (Visscher et al. 1996).

Orthologous mapping and candidate gene identification

Comparative mapping was done by blasting the primer sequence of the markers enclosing QTL regions against the chicken genome sequences on the Sanger Institute website (http://www.ensembl.org) using the basic local alignment search tool (BLASTN). A score is given for matching and mismatching nucleotides and gaps. The total score is given by obtaining the sum of all matches, mismatches and gap penalties for sequence. The E-value or expect score is the number of different values that are equivalent to or better than the score that are expected to occur in a database by chance. A score of more

than 45, percentage identity of greater than 70%, and an E-value less than 0.05 are considered to be significant (Pertsemlidis and Fondon, 2001; Jiang and Michal, 2003). Each matching sequence was then compared with the mouse and human genome sequences to identify regions of homology. Information on genes at or around the QTL location on the chicken genome and their respective homologs in humans and mouse were obtained from http://www.ensembl.org by identifying genes within homologous regions.

REFERENCES

Abasht B., Pitel F., Lagarrigue S., Le Bihan-Duval E., Le Roy P., Demeure O., Simon J., Cogburn L., Aggrey S., Vignal A., and Douaire M., 2006. Fatness QTL on chicken chromosome 5 and interaction with sex. *Genet. Sel. Evol*. (In Press).

Aggrey, S.E., and Cheng, K.M., 1994. Animal model analysis of genetic (co)variance of growth traits in Japanese quail. *Poult. Sci*. **73**:1822-1828.

Aggrey, S.E., Carre, W., Wang, X., Pitel, F., Vignal, A., Le Bihan-Duval, E., Beaumont, C., Duclos, M., Porter, T.E. Simon, J., and Cogburn, L.A., 2005. Mapping of quantitative trait loci for fatness and breast meat yield in a novel resource population of broiler chickens. Proc. Plant & Animal Genomes XIII conf. Jan 15-19, 2005, San Diego, CA http://www.intl-pag.org/13/abstracts/PAG13_P553.html

Ankra-Badu, G.A., and Aggrey, S.E., 2005 Identification of candidate genes at quantitative trait loci on chicken Chromosome Z using orthologous comparison of chicken, mouse, and human genomes. *In Silico Biol*. (In Press)

Borja, A.Z.M., Meijers, C., and Zeller, R. 1993. Expression of alternative spliced bFGF first coding exons and antisense mRNA during chicken embryogenesis. *Dev. Biol*. **157**:110-118.

Boss, O., Hagen, T, and Lowell, B.B. 2000. Uncoupling proteins 2 and 3: potential regulators of mitrochondrial energy metabolism. *Diabetes* **49**:143-156.

Burt, D. and Hocking, P.M. 2002. Mapping quantitative trait loci and identification of genes that control fatness in poultry. *Proc. Nutr. Soc*. **61**:441-446.

Cahaner, A., and Nitsan, Z. 1985. Evaluation of simultaneous selection for live body weight and against abdominal fat in broilers. *Poult. Sci*. **64**:1257-1263.

Churchill, G.A., and Doerge, R.W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138**:963-971.

Comuzzie, A.G. and Allison D.B. 1998. The search for human obesity genes. *Science* **280**: 1374-1377.

Darvasi, A. 1998. Experimental strategies for the genetic dissection of complex traits in animal models. *Nat. Genet*. **18**:19-24.

Davis, W.L., Kipnis, M., Shibata, K., Farmer, G.R., Cortinas, E., Mathews, J.L., and Goodman, D.B. 1989. The immunohistochemical localization of superoxide dismutase activityin the avian epithelial growth plate. *Histochem. J*. **21**:210-215.

De Koning, D.J., Haley, C.S., Windsor, D., Hocking, P.M., Griffin, H., Morris, A., Vincent, J., and Burt, D.W. 2004. Segregation of QTL for production traits in commercial meat-type chickens. *Genet. Res*. **83**:211-220.

Dunne, J., Hanby, A.M., Poulsom, R., Jones, T.A., Sheer, D., Chin, W.G., Da, S.M., Zhao, Q., Beverly, P.C.L., and Owen, M.J. 1995. Molecular cloning and tissue expression of *FAT*, the human homologue of the Drosophila *fat* gene that is located on chromosome 4q34-q35 and encodes a putative adhesion molecule. *Genomics* **30**:207-223.

Erlanson-Albertsson, C., 2003. The role of uncoupling proteins in the regulation of metabolism. *Acta Physiol Scand*. **178**:405-412.

Evock-Clover, C., Poch, S., Richards, M. Ashwell, C., and McMurtry, J. 2002. Expression of an uncoupling protein gene homolog in chickens. Comp. Biochem. Physiol. A Mol. Inegr. Physiol. 133: 345-358.

extensive linkage conservation-but with gene order arrangements between the pig and the human genomes. *Genomics* **25:** 682-690.

Geissler, E.N., Ryan, M.A. and Housman, D.E. 1988. The dominant white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. *Cell* **55**:185-192.

genome regions by identifying their human orthologs. Genome **46,** 798-808.

Gibson, G. and Mackay, T.F.C. 2002. Enabling population and quantitative genomics. *Genet. Research* **80**: 1-6.

Green, P., Falls, K. and Crooks, S. 1990. Documentation for CRI-MAP version 2.4. Washington University School of Medicine, St. Louis, MO.

Haley, C.S., Knott, S.A., and Elsin, J.M. 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**:1195-1207.

Havenstein, G.B., Ferket, P.R., and Qureshi, M.A. 2003. Carcass composition and yield of 1957 versus 2001 broiler when fed representative 1957 and 2001 broiler diets. *Poult. Sci*. **82**:1509-1518.

Hirona, S., Suzuki, S.T., and Redies, C.M. 2003. The cadherin superfamily in neural development: diversity, function and interaction with other molecules. *Front. Biosci.* **8**:D306-D355.

Iannucelli, N., Wolosyn, N., Arhainx, J., Gellin, J. and Milan, D. 1996. GEMMA: A database to automate microsatellite genotyping. *Anim. Genet*. **27** (Suppl. 2): 55.

Ikeobi, C.O.N., Wooliams, J.A., Morrice, D.R., Law, A., Windsor, D., Burt, D.W., and Hocking, P.M. 2002. Quantitative trait loci affecting fatness in the chicken. *Anim. Genet*. **33**:428-425.

Ikeobi, C.O.N., Wooliams, J.A., Morrice, D.R., Law, A., Windsor, D., Burt, D.W., and Hocking, P.M., 2004. Quantitative trait loci for meat yield and muscle distribution in broiler layer cross. *Livest. Prod. Sci*. **87**: 143-151.

International Chicken Genome Sequencing Consortium 2004 Sequence and Comparative Analysis of the Chicken Genome provide unique perspective of vertebrate evolution. *Nature* **432**:695-716.

Jennen D. G., Vereijken A. L., Bovenhuis H., Crooijmans R. P., Veenendaal A., van der Poel J. J., Groenen M. A., 2004. Detection and localization of quantitative trait loci affecting fatness in broilers. *Poult. Sci*. **83**: 295-301.

Jerez-Timaure, N.C., Kearney, F., Simpson, E.B., Eisen, E. J., and Pomp, D. 2004. Characterization of QTL with major effects on fatness and growth on mouse chromosome 2. Obesity Res. 12: 1408-1420.

Jiang, Z. and Michal, J.J. 2003. Linking porcine microsatellite markers to known

Johansson, M., Ellegren, H. and Andersson, L. 1995. Comparative mapping reveals

Jump, D.B., and Oppenheimer, J.H., 1985. High basal expression and 3,5,3'-triiodothyronine regulation of messenger ribonucleic acid S14 in lipogneic tissues. Endocrinology 117:2259-2266.

Kardami, E., and Fandrich, R.R. 1989. Basic fibroblast growth factor ini atria and ventricles of the vertebrate heart. *J. Cell Biol*. **109**:1865-1875.

Knott, S.A., Markund, L., Haley, C.S. Andersson, L. 1998. Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and Large White pigs. *Genetics* **149**:1069-1080.

Lagarrigue S., Pitel F., Carré W., Abasht B., Le Roy P., Neau A., Amigues Y., SourdiouxM., Simon J., Cogburn L., Aggrey, S., Leclercq B., Vignal A., and Douaire M. 2006. Mapping quantitative trait loci affecting fatness and breast muscle weight in experimental meat-type chicken lines divergently selected on fatness. *Genet. Sel. Evol*. (In press).

Le Bihan-Duval, E, Berri, C., Baéza, E. Millet, N., and Beaumont, C. 2001. Estimation of the genetic parameters of meat characteristics and their genetic correlations with growth and body composition in an experimental broiler line. Poult. Sci. 80:839-843.

Lionikas, A., Blizard, D.A., Vandenbergh, D.J., Glover, M.G., Stout, J.T., Vogler, G.P., McClearn, G.E., and Larsson, L. 2003. Genetic architecture of fast- and slow-twitch skeletal muscle weight in 200-day-old mice of the C57BL/6J and DBA/2J lineage. *Physiol. Genomics 16*:141-152.

Mahoney, P.A. Weber, U. Onofrechuk, Biessmann, H. Bryant, P.J., and Goodman, C.S. 1991. The fat tumor suppressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily. *Cell* **67**:853-868.

Matsumoto, H., Silverton, S.F., Debolt, K. and Shapiro, I.M. 1991. Superoxide dismutase and catalase activities in the growth cartilage: relationship between oxidoreductase activity and chondrocyte maturation. *J. Bone Miner. Res*. **6**:569-574.

McElroy, J.P., Harry, D.E., Dekkers, J.C.M., and Lamont, S.J., 2002. Molecular markers associated with growth and carcass traits in meat-type chickens. Vol 30: Pg 209-

212 In Proceeding of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France.

Nuzhdin, S.V., Pasyukova, G., Dilda, C.L., Zhao-Bang, Z. and Mackay, T.F.C. 1997. Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster. *Proc. Natl Acad. Sci. USA* **94**:9734-9739.

Pasyukova, E.G., Vieira, C., and Mackay, T.F 2000. Deficiency mapping of quantitative trait loci affecting longevity in Drosophila melanogaster. *Genetics* **156**: 1129-1146.

Persemlidis A. and Fondon J.W. 2001. Having a BLAST with bioinformatics (and avoiding BLASTphemy). *Genome Biol.* **2:** 2002.1-2002.10.

Pomies, P., Macalma T., and Meckerle M.C. 1999. Purification and characterization of an alpha-actinin-binding PDZ-LIM protein that is up-regulated during muscle differentiation. *J. Biol. Chem*. **274**:29242-29250.

Pomp, D. 1997. Genetic dissection of obesity in polygenic animal models. *Biochem. Genet*. **27**: 285-306.

Remington, D.L., Ungere, M. and Purugganan, M.D. 2001. Map-based cloning of quantitative trait loci: progress and prospects. *Genet. Research* **78**:213-218.

Ricard, F.H., 1975. Essai de selection sur la forme de la courbe de croissance chez le poulet. *Ann. Genet. Sel. Anim*. **7**:427-443.

Ricard, F.H., and Rouvier, R. 1967. Étude de la composition anatomique du poulet. I. Variabilité de la repartition des differentes parties corporelles chez des coquelets “Bresse-Pile”. *Ann. Zootech*. **16**:23-29.

Sasaki, O., Odawara, S., Takahasi, H., Nirasawa, K., Oyamada, Y., Yamamoto, R., Ishii, K., Nagamine, Y., Takeda, H., Kobayashi, E., and Furukawa, T. 2004. Genetic

mapping of quantitative trait loci affecting body weight, egg character and egg production in F2 intercross chicken. *Anim. Genet*. **35**:188-194.

Schmid M., Nandra, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde J.M., Wimmers, K., Burt, D.W., Smith, J., A'Hara, S., Law, A., Griffin, D.K., Bumstead, N., Kaufman, J., Thompson, P.A., Burke, T., Groenen, M.A., Crooijmans, R.P., Vignal, A., Fillon, V., Morisson, M., Pietel, F., Tixier-Boichard, M., Ladjali-Mohammedi, K., Hillel, J., Maki-Tanila, A., Cheng, H.H., Delany, M.E., Burnside, J., and Mizuno, S. 2000. First report on chicken genes and chromosomes 2000. *Cytogenet. and Cell Genet*. **90**:169-218.

Schűtz, K., Kerje, S., Carlborg, O., Jacobsson, L., Andersson, L. 2002. QTL analysis of a Red Junglefowl x White Leghorn intercross reveals trade-off in resource allocation between behavior and production traits. *Behav. Genet*. **32**:423-433.

Seaton, G., Haley, C.S., Knott, S.A., Kearsey, M., and Visscher, P.M. 2002. QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics* **18**:339-340.

Sewalem, A., Morrice, D.M., Law, A., Windsor, D., Haley, C.S., Ikeobi, C.O.N., Burt, D.W., and Hocking, P.M. 2002. Mapping a quantitative trait loci for body weight and three, six and nine weeks of age in a broiler layer cross. *Poult. Sci*. **81**:1775-1781.

Sumitomo, S., Saito, T. and Nohno, T. 1993. A new receptor protein kinase from chick embryo related to type II receptor for TGF-beta. DNA seq. 3:297-320.

Tsuiskula-Haavisto, M., Honkatukia, M., Vilkki, J., de Koning, D.-J., Schulman, N.F., and Maki-Tanila, A. 2002. Mapping of quantitative trait loci affecting quality and production traits in egg layers. *Poult. Sci*. **81**:919-927.

Van Kaam, J.B.C.M.H., van Arendonk, J.A.M., Groenen, M.A.M., Bovenhius, H., Vereijken, A.L.J., Crooijmans, R., van der Poel, J.J., and Veenendaal, A. 1998. Whole genome scan for quantitative trait loci affecting body weight in chickens using a three generation design. *Livest. Prod. Sci*. **54**:133-150.

Visscher, P.M., Thompson, R., and Haley, C.S. 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**:1013-1020.

Wang, X., Carre, W., Reijto, L. and Cogburn, L.A., 2002. Global gene expression profiling in liver of thyroid manipulated and/or growth hormone injected broiler chickens. *Poult. Sci*. **81** *(Suppl. 1)*:63.

Wang, X., Carre, W., Zhou, H., Lamont, S.J., and Cogburn, L.A., 2004. Duplicated Spot 14 genes in the chicken: characterization and identification of polymorphisms associated with abdominal fat traits. *Gene* **332**:79-88.

Xia, H., Winokur, S.T., Kuo, W.-L, Altherr, M.R., and Bredt, D.S., 1997. Actin associated LIM protein: Identification of a domain interaction between PDZ and spretrin-like repeat motifs. J.Cell Biol. 139:507-515.

Zerehdaran, S., Vereijken, A.L.J., van Arendonk, J.A.M., and van der Waaij, E.H. 2004. Estimation of genetic parameters for fat deposition and carcass traits in broilers. *Poult. Sci*. **83**:521-525.

Table 1 Number of markers, map length and first and last markers for each chromosome

Chromosome	Number of markers	Map length (cM)	First Marker	Last Marker
1	18	558	MCW0168	MCW0108
2	17	432	MCW0205	MCW0157
3	11	289	ADL0177	MCW0037
4	6	233	ADL0317	LEI0037
5	8	141	LEI0082	ADL0298
6	4	93	LEI0192	ADL0323
7	6	151	ADL0315	LEI0064
8	3	54	LEI0136	MCW0305
9	3	71	LEI0028	ADL0132
10	4	53	LEI0112	MCW0067
11	3	49	MCW0097	ADL0308
12	3	31	ADL0372	ADL0044
13	3	35	MCW0110	MCW0213
14	2	43	MCW0123	MCW0123
15	3	50	MCW0211	ADL0206
17	1	21	ADL0293	ADL0199
18	2	17	ADL0304	MCW0217
19	2	15	MCW0266	MCW0256
26	3	29	ADL0330	LEI0074
27	3	13	MCW0146	MCW0233
Z	3	136	ADL0117	MCW0128

Table 2. Statistical tests (F-ratio), QTL position, flanking markers and 95% confidence interval for carcass traits in an F_2 population derived from a divergent chicken line selected for fast or slow growth

Chromosome	F-ratio	Position cM	Flanking markers	95% Confidence interval
Breast meat weight, g				
1	5.36^{+}	240	LEI101-MCW068	19 - 529
2	6.47^{+}	342	MCW264-MCW056	24 - 368
3	4.81^{+}	176	MCW083-MCW150	0 - 280
4	22.64**	205	MCW240-LEI073	184 - 216
Breast meat yield, %				
1	9.71*	221	ADL252-LEI101	145 - 506
3	8.78*	68	MCW150-LEI032	21 - 90
6	8.97*	115		
Abdominal fat weight, g				
1	8.28*	533	LEI061-ADL328	0 - 539
4	6.17^{+}	203	MCW240-LEI073	
5	9.28*	40	MCW193-MCW214	9 - 90
Abdominal fat yield, %				
1	10.71*	534	LEI061-ADL328	0 - 539
4	15.35**	204	MCW240-LEI073	120 - 221
5	8.72*	40	MCW193-MCW214	12 - 92
Thigh weight, g				
2	7.68*	347	MCW264-MCW056	29 - 376
4	37.30**	209	MCW240-LEI073	201 - 219
5	7.49*	43	MCW193-MCW214	
26	7.66*	11	ADL330-MCW069	0 - 31
P. major weight, g				
2	5.97^{+}	340	MCW264-MCW056	29 - 440
3	5.17^{+}	174	MCW083-MCW150	0 - 249
4	20.41**	206	MCW240-LEI073	185 - 216
10	7.28	9	ADL209-MCW067	0 - 62
P. major yield, %				
1	9.85*	237	ADL328-ADL328	74 - 506
3	8.77*	69	MCW150-LEI032	33 - 83
6	7.80*	115		

P. minor weight, g				
1	6.13[+]	242	LEI101-MCW068	9 - 528
2	10.09**	338	MCW264-MCW056	86 - 366
4	24.05**	204	MCW240-LEI073	189 - 215

[+]significant at chromosome wide level at P≤0.05
*significant at genome-wide level at P≤0.05
**significant at genome-wide level at P≤0.01

Table 3 Mean (± SE) of additive effect, effects as a proportion of the phenotypic standard deviation (SD) and the proportion of variance explained by QTL in males and females

	Male		Female		
Chromosome	Additive effect	SD	Additive effect	SD	% Variance
Breast meat weight, g					
4	4.14± 0.80	0.50	4.06± 0.94	0.49	6.44
Breast meat yield, %					
1	0.06± 0.04	0.16	0.18±0.04	0.43	2.67
3	0.10± 0.04	0.25	0.14±0.04	0.34	2.36
6	-0.03± 0.04	0.07	0.20±0.05	0.50	2.42
Abdominal fat weight, g					
1	0.81± 0.94	0.08	4.09±1.03	0.42	2.26
5	3.02± 0.97	0.31	3.44±1.12	0.36	2.56
Abdominal fat yield, %					
1	0.05± 0.07	0.07	0.36±0.07	0.48	3.00
4	-0.35± 0.07	-0.47	-0.23±0.08	-0.31	4.37
5	0.21± 0.07	0.27	0.28±0.08	0.38	2.40
Thigh weight, g					
2	5.93±1.86	0.37	4.31±1.87	0.27	2.08
4	10.30±1.61	0.64	10.89±1.86	0.67	10.40
5	3.73±1.57	0.23	5.68±1.81	0.34	2.02
26	2.60±1.42	0.16	5.67±1.64	0.35	2.07
P. major weight, g					
4	3.07±0.06	0.50	2.85±0.71	0.46	5.81
P. major yield, %					
1	0.03±0.02	0.09	0.13±0.03	0.40	2.75
3	0.08±0.03	0.25	0.11±0.03	0.35	2.37
6	-0.02±0.04	-0.08	0.15±0.03	0.46	2.09
P. minor weight, g					
2	0.94±0.28	0.41	0.86±0.28	0.38	2.81
4	1.09±0.22	0.48	1.24±0.26	0.54	6.80

1
2

Table 4 QTL by sex interaction for carcass traits in an F_2 population derived from a divergent chicken line selected for fast or slow growth

Trait	F-ratio	Location (cM)[1]	Additive effect	% variance[2]
GGA1				
Breast meat yield, %				
Male	4.52	212 (62-479)	-0.11±0.05	1.11
Female	12.64*	239 (48-505)	0.15±0.04	4.20
Fat weight, g				
Male	4.66	142 (26-514)	2.34±1.08	1.18
Female	17.98*	539 (14-542)	3.93±0.93	6.00
Fat yield, %				
Male	3.02	141 (83-528)	0.13±0.08	0.65
Female	20.24*	541 (12-542)	0.34±0.08	6.72
P. major yield, %				
Male	3.96	221 (68-434)	-0.07±0.04	0.88
Female	14.55*	240 (218-506)	0.13±0.03	4.80
GGA3				
Breast weight, g				
Male	11.48*	172 (7-277)	4.63±1.37	3.27
Female	3.92	287 (7-287)	-1.72±0.87	1.08
P major weight, g				
Male	12.13*	171 (185-280)	3.58±1.03	25.79
Female	3.56	287 (21-287)	-1.24±0.06	0.95
GGA26				
Thigh weight, g				
Male	4.82	20 (0-41)	4.04±1.84	1.22
Female	12.51*	11 (2-25)	5.62±1.60	4.13

*Significant at genome-wide level at $P \leq 0.05$

[1]95% confidence interval in brackets

[2] Proportion of the phenotypic variance explained by QTL

Table 5 Candidate genes derived from differentially expressed trancriptome and known genes around QTL location (209 cM, flanking markers MCW240-LEI073) on chicken GGA4 and their respective homologues in mouse and human

	Differentially expressed cDNA clones		Putative candidate genes		
Trait[1]	Clone	Gene name	GGA	MMU	HSA
BMW	pgp1n.pk013.n16	Annexin A5 (Annexin V) (Lipocortin V) (Calphobindin I) (CBP-I)	Phox 2b protein	Quininoid dihydropterine reductase	Slit 2 homolog 2 protein precursor
AFP			Neuronal acetylcholine receptor subunit alpha 4 precursor	Cytosol aminopeptidase	Potassium channel interacting protein 4 isoform 3
TW				Endothelial derived gene 1	
PMaW	pgl1c.pk002.h11	Alcohol dehydrogenase *ADH-F*		Mblk related protein 1	
PMiW	pgf2n.pk005.m5	Ubiquitin-conjugating enzyme E2D 3 (*UBE2D3*)	Type IIb sodium phosphate cotransporter	Slit homolog 2 protein precursor	Probable G-protein coupled receptor 125 precursor
	pgm2n.pk006.b11	FAT tumor suppressor (Drosophila) homolog [Rattus norvegicus]	Kv channel interacting protein	Potassium channel interacting protein	Cysolic beta glucosidase
	pgf1n.pk010.p14	Multifunctional protein ADE2 (SAICAR synthetase) (AIR carboxylase)	Slit 2		Peroxisome proliferator activated receptor gamma coactivator 1 alpha
	pgl1n.pk001.e1	Amidophosphoribosyltransferase (ATASE)	FGF binding protein	Peroxisome proliferator activated receptor gamma coactivator 1 alpha	Peroxisome proliferator activated receptor gamma coactivator 1 alpha
	pgf2n.pk006.a17	C1q and tumor necrosis factor related protein 7 (*C1QTNF7*)	LIM domain binding protein CLIM 1		
	pgf2n.pk004.k13	Spondin 2 *(SPON2),* extracellular matrix protein	WD repeat protein	Putative pre-mRNA splicing factor RNA helicase	Putative pre-mRNA splicing factor RNA helicase
	pgp1n.pk003.o19	Superoxide dismutase 3, r (*EC-SOD*)	Homeobox protein		
			Dopamine receptor D1B	Extracellular superoxide dismutase	Soluble liver antigen
			Collapsin response mediator protein 1B		Extracellular Superoxide dismutase
			Lz1 mRNA	Solute carrier family 34	
			Fibroblast growth factor 2	Leucine rich repeat LG1 family member 2 precursor	Recombining binding protein suppressor of hairless (J kappa signal recombination binding protein)
			Bone morphogenetic protein receptor, type IB (receptor protein kinase)	Phosphatidylinositol 4 kinase type II beta	
			Fat tumor suppressor homolog	Anaphase promoting complex subunit 4	
			Melatonin receptor 1A	Recombining binding protein suppressor of hairless (J kappa signal	Cholecystokinin tye A receptor
			Platelet derived growth		

	factor receptor Tec protein tyrosine kinase Ankyrin 2 Fibroblast growth factor receptor 3 Fibroblast growth factor receptor-like 1 Regulator of G-protein signaling 12 v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog Actin filament associated protein Carboxypeptidase Z Protein kinase Alcohol dehydrogenase 1	recombination binding protein) Cholecystokinin type A receptor (Gene)Similar to stromal interaction molecule 2	Similar to stromal interaction molecule 2 Phosphatidylinositol 4 kinase type II beta Leucine rich repeat LG1 family member 2 precursor TBC 1 domain family member 19 Zinc finger CCHC domain containing protein 4 Predicted hypothetical protein XP-371691 Anaphase promoting complex subunit 4 Solute carrier family 34

[1]BMW=Breast meat weight; AFP=Abdominal fat percentage; PMaW=*Pectoralis major* weight; PMiW= *Pectoralis minor* weight.

CHAPTER 6

THE GENOMIC LANDSAPE OF A MAJOR QTL REGION ON CHICKEN CHROMOSOME 4: CpG ISLANDS, GENE DENSITY AND REPEATS[1]

[1] Georgina A. Ankra-Badu and Samuel E. Aggrey. To be submitted to Genomics

ABSTRACT

This study was conducted to characterize the genomic landscape of a major QTL region on chicken chromosome 4 (68-80 Mb) and to identify novel candidate genes by CpG island detection and comparative mapping. One hundred and nine known genes and 179 CpG islands were detected at this locus and thirty four percent of these genes contained CpG islands. The segment spanning 68-70 Mb had the highest CpG island density while that spanning 78-80 Mb contained the highest number of genes. Analysis of the distribution of repetitive elements showed that LINE, low complexity and simple repeats constituted the majority of repetitive DNA in the QTL region. Generally regions with a high GC content and gene and CpG island density had a relatively low percentage of repeats. Comparative mapping identified a total of seventy three genes from a match with twenty species. These genes are involved in protein synthesis, transcriptional regulation and several other functions. Six probable novel genes were identified on GGA4 by blasting these genes against the chicken genome. Three of these genes are housekeeping genes which are either involved in protein transport or signal transduction. A putative ortholog of rhotekin, which is a housekeeping gene with nucleotide binding and apoptotic properties, may be linked with the acetylcholine receptor.

INTRODUCTION

Most traits of agricultural importance have a complex mode of inheritance and are controlled by several gene loci, also referred to as Quantitative trait loci (QTL). Detection of the individual genes underlying these traits is essential but difficult since these QTL regions span very large regions of the chromosome and are influenced by the environment. Some successful methods that have been developed to isolate these genes

include fine mapping (Talbot et al., 1999), comparative mapping (Yu et al., 1997; Ankra-Badu and Aggrey, 2005), candidate gene analysis (Hill, 1998) and the identification of CpG islands (Gillespie et al., 1991; Cross et al., 2000). The first three methods though very useful for the detection of novel genes have limitations which may be overcome by the location of CpG islands. Candidate gene analysis for example, is only able to detect very strong associations between the gene and the trait due to sample size limitations (Hill, 1998). This kind of study may also be prone to false positives because differences between mutations and rare chance polymorphisms are usually indistinguishable (Macrae, 2000). The limitations of comparative mapping are associated with modifications in our understanding of gene homology (Wakefield, 1998). The distinctive nature of CpG islands makes them a powerful additional tool for the detection of genes that control complex traits.

CpG islands are short stretches of unmethylated DNA with a higher than average frequency of CpG nucleotides (Larsen et al. 1992). They are defined as a region with at least 200 base pairs (bp), a GC percentage greater than 50% and an observed to expected CpG ratio of greater than 0.6 (Gardiner-Garden and Frommer, 1987). A more stringent characterization to exclude Alu repeats classifies a CpG island as a region of least 500 bp, a GC% greater than 55 and an observed to expected ratio of at least 0.55 (Takai and Jones, 2002). CpG islands are usually associated with the 5` and 3` ends of genes, and are associated with about 40% of promoters in mammalian genes. This implies that at least half of all genes in mammalian genomes are linked to CpG islands (Gardiner-Garden and Frommer, 1987; Larsen et al., 1992; Antequera and Bird, 1993; Wang and Leung, 2004). CpG islands in the promoter regions of genes of normal (non cancerous) tissues are

unmethylated if the genes are expressed hence the methylation of CpG sites in the promoter region determines whether a gene will be expressed or not. The methylation of CpG islands is also associated with genomic imprinting, X-chromosome inactivation and cancer (Feil and Khosla, 1999; Panning and Jaenisch, 1998 and Yoder et al., 1997).

The detection of CpG islands for the genetic analysis of QTL regions and the identification of candidate genes has been reported in literature. Weber et al. (1991) identified several CpG islands in a region associated with Huntington's disease and reported that sequences associated with CpG islands reveal cross-species conservation. In another study, Lee et al. (2006) found a total of eight genes and 33 CpG islands in a QTL region for fat and body weight on pig chromosome 6. They indicated that the majority of these islands were located close to the putative candidate genes in the QTL location. Gillespie et al. (1991) identified candidate genes for autosomal dominant polycystic kidney disease by locating CpG islands in a region associated with the disease. The CpG islands were used as markers to search for genes at this locus which is known to have a high density of these islands. CpG islands have also been used to estimate the number and expression pattern of genes. Larsen et al. (1992) reported that CpG islands are located at the 5' ends of all housekeeping genes compared to only 40% of genes with restricted expression. Antequera and Bird (1993) estimated the number of genes on the human and mouse genomes by determining the total number of CpG islands associated with genes on these genomes.

Chicken chromosome 4 (GGA4) has been identified as an important location for growth, body weight, body composition and skeletal and egg traits (Tuiskula-Haavisto et al., 2002; Schreiweis et al., 2005). Six thousand six hundred and sixty two CpG islands

(ww.ncbi.nlm.nih.gov) and 1288 known and probable genes (http://www.ensemble.org) have been identified on this chromosome. According to Hillier et al. (2004) nearly half of all CpG islands on the chicken genome are associated with genes. Therefore detailed mapping of CpG islands could aid in the identification and prediction of additional putative novel genes in QTL regions (Cross and Bird, 1995). CpG islands are usually under represented in the genome due to the methylation of the cytosine nucleotide (Bird, 1987). Alu repeats have been found to protect these islands from *de novo* methylation which leads to gene silencing and the displacement of transcription factors (Kang et al., 2006; Fazzari and Greally, 2004). Therefore, it is equally important to study the relationship between CpG islands and repetitive sequences. The objective of this study was to characterize a major QTL region (68-80Mb) on GGA4 through the identification of CpG islands, putative novel genes and repeats.

MATERIALS AND METHODS

Genes in the QTL location for growth were extracted from BioMart in the ensembl web server (http://www.ensembl.org). Novel genes and those with no description were categorized as unknown genes. CpG islands in these genes were located by the cpgplot and cpgreport programs in the EMBOSS package (http://emboss.sourceforge.net) via the EMBL-EBI web server (www.ebi.ac.uk). The criteria for a CpG island were by Gardiner and Frommer (1987). These islands are defined as a sequence of at least 200 bp with a GC content of at least 50% and an observed to expected ratio of at least 0.6. The ratio of observed to expected CpG islands was calculated as the number of CpG × N / total number of Cytosine × total number of

Guanine nucleotides (Gardiner-Garden and Frommer, 1987) and were estimated by using a 100 bp window along the sequence at 1bp intervals, where N is the total number of nucleotides in the sequence. Information on CpG islands in the QTL region was also obtained from the EMBL-EBI web server (www.ebi.ac.uk) and the analysis was conducted in 2-Mb segments due to the large size of the region (~12 Mb). Calculation of gene and CpG island density in the region was done by dividing the total number of CpG islands and genes by the total length in mega bases. In order to identify novel candidate genes in gene-poor regions, BLAST analysis was conducted on segments in the QTL region that contained CpG islands but no known genes. These areas were located by comparing the location of CpG islands with the positions of all genes in the QTL region. The gene-poor sections were blasted against the genomes of 23 species to identify homologous regions via the Sanger Institute website (www.ensembl.org) A score of more than 45, a percentage identity of greater than 70%, and an e-value less than 0.05 were considered to be significant (Pertsemlidis and Fondon, 2001; Jiang and Michal, 2003). Putative candidate genes were extracted from these regions using Biomart (www.ensembl.org). These were then blasted against the chicken genome to identify orthologs and their positions if they were present. The best matches were classified based on the criteria outlined above.

Repetitive sequences (LINE, SINE, LTR, simple and low complexity repeats) in the QTL region were detected by using Repeat Masker (Smit, 2004) and a summary of information on the repeats was obtained using the table format in the UCSC genome browser (Karolchik, 2004).

RESULTS

There were 179 CpG islands (www.ebi.ac.uk\emboss/cpgplot) and 6,587 repeats in the major QTL region (Table 1). The total number of repeats in each category is shown on Table 1. There were 2,072 LINE and 2,020 complexity repeats in this region. These two groups represented about 62% of the total number of repeats. Simple repeats made up about 25% and there were 372 DNA repeats and 472 LTR repeats. Satellite repeats were the least and comprised of less than 1% of the total. Table 2 shows the genomic landscape in 2-Mb segments. The areas spanning 70-72 Mb and 72-74 Mb had the highest percentage of both low complexity and LINE repeats. The gene to Cpg island ratio for all segments was either 1:2 or 1:1 except for the region covering 74-76 Mb which showed a gene to CpG island ratio of 2:1. The overall CpG island to gene ratio was approximately 1:0.8.

The list of genes in the QTL region is summarized in Table 3. A total of 136 comprising of 109 known and 27 unknown genes were identified in this region which spanned approximately 12Mb. There was an average of 11 genes per Mb and gene density was highest between 68 to 70 Mb and 78-80 Mb, 18 and 19 genes per Mb respectively. Forty seven genes (about 34% of the total) which included four unknown genes contained CpG islands. The shortest single CpG island was 200 bp in length while the longest which was associated with a hypothetical protein was 989 bp. The coiled coil domain containing 4 (CCDC4) gene, which covered 166,585 bp, had the highest number of CpG islands. This gene had 6 CpG islands with an average length of 674 bp. Appendix A shows the genes identified from comparative mapping of the QTL region for growth on GGA4 with twenty three other species. The region spanning 68-70 Mb matched with 7

species which contained genes. Some of the genes identified through comparative mapping in region 68-70 Mb were rhotekin from *Takifugu rubripes*, BAZ1_A from *Macaca mulatta* and a gene from the UDP transporter solute carrier family 35 family on *Dasypus novemcinctus*. The region from 70-72 Mb only identified genes from *Cioni intestinalis* and *Drosophila melanogaster* and some of these were innexin shaking B and frequenin-1. Fourteen regions matched with the QTL region from 72-74 Mb and the most common gene was the Protocadherin precursor gene. Other genes were the circadian pacemaker protein and carboxylase precursor gene from *Xenopus tropicalis* and *Monodelphis domestica* respectively. A gene from the Plaur domain containing 3 precursor family and another from the insulin gene enhancer islet family were identified in *Dasypus novemcinctus* from comparative mapping of the 74-76 Mb segment. The segment covering 76-78 Mb matched 18 species and corresponded to genes like chemokine receptor type 5 antigen from *Oryctolagus cuniculus* and the EF – hand domain containing 2 from *Mus musculus*. The last segment (78- 80 Mb) matched with 17 species. The MIDN gene from *Drosophila melanogaster* and the histone H2A gene from *Ciona intestinalis* were located in the respective corresponding regions. The genes identified from the first comparative mapping were blasted against the chicken genome to determine the position of their orthologs on the chicken chromosome. Six genes matched with homologous positions on GGA4 (Table 4). Rhotekin for example was located in the region of GGA4 that matched with scaffold_124:706692-710710 in *Takifugu rubripes* from the first BLAST analysis. The CG4080-PA gene was also found in the GGA4 region corresponding to position 3L: 9360491-9362517 in *Drosophila melanogaster*. The

GGA4 match for the TIM44 gene did not meet the criteria for the e-value but was selected based on a good score and a high percentage identity (96%).

DISCUSSION

The CpG island to gene ratio is GGA4 is estimated at about 3:1 on the q arm on and 4:1 on the p arm. Furthermore, about 48% of the islands on the chicken genome overlap a gene (Hillier et al. 2004). Contrary to these findings, the overall ratio of CpG islands to genes in the QTL region was approximately 1: 0.8. Additionally, 34% of these genes contained CpG islands (Tables 2 & 3). A more detailed study of the QTL region in 2-Mb segments revealed a CpG island to gene ratio of either 1:1 or 2:1. These high CpG island to gene ratios could be due to the fact that the region of interest is associated with a major QTL for several important traits.

Cross et al. (2000) constructed CpG libraries from HSA 18 and 22. They found that all gene-rich chromosomes were CpG rich and the reverse was true for gene-poor chromosomes. They therefore concluded that the density of CpG islands could be used to predict the number of genes in a chromosomal region. Antequera and Bird (1993) in an earlier study, predicted the number of genes on the mouse and human chromosomes from the amount of CpG islands on the genomes of these species. The results from our study did not show a very clear relationship between CpG island and gene density because the 2-Mb segment with the highest number of CpG islands did not have the highest number of genes. On the other hand, the segment with the highest number of genes contained only 36 islands. The inconsistency in these results may be attributed to recombination rate (Webster et al., 2006) and other unknown factors since the GC content and number

of repeats were virtually the same for these two regions. However, the segment spanning 70-72 Mb had the least number of genes and the lowest concentration of CpG islands.

Generally, the areas with the highest percentage of repeats had the lowest GC content, CpG island density and number of genes (Table 2). This is consistent with the observation that the density of repetitive sequences has a negative correlation with GC content, CpG islands and gene density, but relates positively to the size of the chromosome (Hillier et al., 2004). The International Chicken Genome Sequencing Consortium (ICGSC) report indicates that the number of interspersed repeats in the chicken is relatively small, 9% compared to 40-50% in mammals, due to low transposable activity. These repeats can be categorized into long interspersed nucleotide elements (LINES), short interspersed nucleotide elements (SINES) and long terminal repeat (LTR) transposons (Pevsner, 2003). The interspersed repeats detected in the QTL region in the present study were LINES and LTRs which made up about 62% of the total number of repeats. Over 80% of LINE repeats in the chicken are CR1s which are GC rich, have truncated 5' ends and are equivalent to L1 LINE repeats in humans. No SINE repeats were detected in this study. This is consistent with the report that SINE repeats, which usually constitute the majority of interspersed repeats, have been non-existent in the chicken for more than 50 Myr though they exist in all other vertebrates (Hillier et al., 2004). Studies on the mammalian genome have shown that a low density of L1 (LINE) repeats in GC-rich regions corresponds to a high concentration of genes and CpG islands (Fazzari and Greally, 2004). This is due to the fact that retroelements can cause a spread in CpG methylation which leads to the loss of CpG islands (Kang et al., 2006). Conversely, Alu (SINE) repeats protect CpG islands and are often found in higher

proportions in housekeeping genes than tissue specific genes (Kang et al., 2006). Therefore in humans, all housekeeping genes contain CpG islands compared to 40% of tissue specific genes (Larsen et al., 1992). Due to the absence of Alu repeats in the chicken, it is expected that the chicken genome will have fewer CpG islands than humans. On the contrary, there are 45,000 CpG islands in the human (Antequera and Bird, 1993) compared to 70,000 in the chicken (Hillier et al., 2004). This observation may be explained by the compact nature of the chicken genome which has a small number of repetitive DNA compared to the human genome. Ovcharenko et al. (2005) characterized gene deserts in the human genome and reported that these regions have a very low GC content and a high number of repeats relative to the whole genome. The distribution of these repeats indicated that the number of LINE repeats in this region was much higher than the amount of SINE repeats when compared to the entire genome. The reverse was true for gene-rich regions. Belle et al. (2005) have however indicated that that the distribution of interspersed repeats in AT or GC regions is not homogenous because it depends on age. In this study, GC content did not vary greatly among segments but areas with lower GC content tended to have a higher proportion of LINE repeats.

Thirty four percent of the genes located in the QTL region in this study contained CpG islands. There was an average of one CpG island per gene and an average GC content of 62%. The gene with the highest number of CpG islands was CCDC4 which occupied 166,585 bases and had six CpG islands with an average length of 674 bp. The closest CpG island was located about 350bp away but extended a few mega bases into the gene. Larsen et al. (1992) in their pioneering work on the relationship between tissue

specificity and CpG islands indicated that all housekeeping genes are associated with CpG islands compared to 40% of tissue specific genes. Tissue specific genes without islands are restricted in their expression because they are only available to transcription factors in cells that prevent methylation. Those with CpG islands however, are restricted in their expression by trans-acting repressors in non-expressing tissues (Bird, 1987). This may explain why the CCDC4 gene is restricted in its expression despite its association with numerous CpG islands.

Appendix A summarizes the comparative mapping results between the QTL region on GGA4 and twenty three other species. The comparison was done in 2-Mb segments by blasting gene-poor sections of these segments against the genomes of other species in the ensembl database to identify their putative orthologs. Overall, seventy three genes were identified in 20 species. A number of putative novel genes were identified in the gene-poor regions of the QTL region by blasting these genes against the chicken genome (Table 4). Four out of six genes were involved in protein binding while another gene, rhotekin, was involved in the biological processes of apoptosis and signal transduction. The forkhead box I2 gene is a transcriptional regulator and a cellular component of the nucleus. Rhotekin, the retinoblastoma protein and the forkhead box I2 protein are housekeeping genes thus they are expressed in several tissues. CG4080 is only expressed in the germline and the tissue specificity of the G protein receptor 125 and the TIM44 mitochondrial precursor are not very clear (www.uniprot.org). The location of the CpG island flanking the putative CG4080-PA gene appears to be in the vicinity of a probable 3' end. This is consistent with the observation that while all housekeeping genes in vertebrates are associated with CpG islands at their 5' ends, a number of tissue specific

genes are associated with these islands at their 3' ends (Gardinner-Garden and Frommer, 1997). Larsen et al. (1992) have also reported that though CpG islands are associated with the 5` ends of all housekeeping, their association with tissue specific genes is not restricted to the 5` end of the transcription unit. CpG islands have been detected at the 3' end of the glucose-6-phosphate dehydrogenase gene, yet subsequent studies indicated that those CpG islands were really located at the 5` of two other genes (Bird, 1987). Since there was no known gene in the vicinity of this island it is possible that the CpG island we detected actually occurs at the 5` end of an unknown gene near the probable CG4080-PA gene.

The rhotekin gene which is found in the Fugu, human, frog and other species matched a putative ortholog on the segment spanning 69,430,821-69,434,843 on GGA4. Hillel et al. (2004) reported that the majority of genes conserved in chickens and humans are conserved in fish and may also be preserved in most species. Furthermore, 72% of genes with 1:1 orthologs between chickens and humans have one ortholog in the Fugu, also known as the puffer fish. Moreover, these three species encode a similar set of protein domains. The location of the putative rhotekin gene on GGA4 is no more than 4 x 10^{-3} cM away from the acetylcholine receptor. Coincidentally, the rhotekin gene is involved in acetylcholine receptor activity. It is therefore possible that these genes are linked.

SUMMARY

CpG islands are areas of the genome with higher CpG nucleotide density and GC content than the genome average. They are usually associated with the 5` ends of genes

and are useful for the detection of novel genes due to their unique structure. One hundred and seventy nine CpG islands and 109 genes were identified in the QTL region under study. Despite the fact that only thirty four percent of these genes actually contained at least one CpG island, several other genes were located close to at least one island. Repetitive elements identified in this study showed that the majority are made up of simple, LINE and low complexity repeats. A high CpG and GC content corresponded with a relatively low proportion of repeats while regions with the highest number of CpG islands were associated with a relatively high gene density. Seventy three putative candidate genes in 20 species were identified by comparative mapping analysis of gene-poor regions containing CpG islands. Subsequent BLAST results showed that six genes had probable orthologs in the QTL region on GGA4. One of these genes, rhotekin has been identified in the Fugu (T. rubripes) and is involved in apoptosis and signal transduction (Gene Ontology). The present study has located six putative genes on GGA4 by CpG island analysis and comparative mapping. The authentication of the identities of these genes however requires wet lab analysis and other methods of verification.

REFERENCES

Ankra-Badu, G.A and Aggrey, S.E. 2005. Identification of candidate genes at quantitative trait loci on chicken chromosome Z using orthologous comparison of chicken, mouse and human genomes. In Silico Biology 5: 593-604.

Antequera, B and A. Bird. 1993. Number of CpG islands and genes in Human and Mouse. Proc Natl Acad Sci 90: 11995-11999.

Bird, A.P. 1987. CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3: 342-347.

Cross, S.H and Bird, A.P. 1995. CpG islands and genes. Curr. Opin.Genet.Dev. 5: 309 - 314

Cross, S.H., Clark, V.H., Simmen, M.W., Bickmore, W.A., Maroon, H., Langford, C.F., Carter, N.P and A.P. Bird. 2000. CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes. Mammalian genome 11: 373-383.

Fazzari, M.J and J.M. Greally. 2004. Epigenomics: Beyond CpG islands. Nature Reviews Genetics 5: 446-455.

Feil, R and S. Khosla. 1999. Genomic imprinting in mammals - an interplay between chromatin and DNA methylation? Trends Genet. 15: 431-435.

Gardinner-Garden M and M. Frommer. 1987. CpG islands in vertebrate genomes. J. Mol. Biol. 196: 261-282.

Gillespie, G.A.J., Somlo, S., Germino, G.G., Weinstat-Saslow, D and Reeders, S.T. 1991. CpG island in the region of an autosomal dominant polycystic kidney disease locus defines the 5' end of a gene encoding a putative proton. Proc Natl Acad Sci 88: 4289-4293.

Hill, A.V.S. (1998). Host genetics of infectious diseases: old and new approaches converge. Emerging infectious diseases 4: 695-697.

Hillier, L.W et al. 2004. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695-716.

Jiang, Z and Michal, J.J. 2003. Linking porcine markers to known genome regions by identifying their human orthologs Genome 46: 798-808.

Kang, M., Rhyu, M-G., Kim, Y-H., Jung, Y-C., Hong, S-J, Cho, C-S and H-S Kim. 2006. The length of CpG islands is associated with the distribution of Alu and L1 elements. Genomics 87: 580-590.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. 2004. The UCSC Table Browser data retrieval tool. Nucl. Acids Res. **32**(Suppl 1), D493-D496.

Larsen, F., Gundersen, G., Lopez, R and H. Prydz. 1992. CpG islands as gene markers in the human genome. Genomics 13: 1095-1107.

Lee, K-T., Park, E-W., Moon, S., Park, H-S., Kim, H-Y., Jang, G-W., Choi, B-H., Chung, H.Y., Lee J-W., Cheong, I-C., Oh, S-J., Kim, H., Suh, D-S and T-H Kim. 2006. Genomic sequence analysis of a potential QTL region for fat trait on pig chromosome 6. Genomics 7: 218- 224.

Macrae, C.A. 2000. Genetics and dilated cardiomyopathy: limitations of candidate gene strategies. European Heart Journal 21: 1817-1819.

Ovcharenko, I., Loots, G.G., Nobgrega, M.A., Hardison, R.C., Miller, W and L. Stubbs. 2005. Evolution and functional classification of vertebrate gene deserts. Genome Research 15: 137-145.

Panning, G and R. Jaenisch. 1998. RNA and the epigenetic regulation of X chromosome inactivation. Cell 93: 305-308.

Pertsemlidis, A and Fondon, J.W. III. 2001. Having a BLAST with bioinformatics (and

avoiding BLASTphemy) Genome Biol 2: reviews2002.1-2002.10.

Pevsner J. Bioinformatics and functional genomics. 2003. John Wiley and Sons. Inc. Hoboken, New Jersey. Pp 629-633.

Schreiweis, M.A., Hester, P.Y., Settar, P and Moody, D.E. 2006 Identification of quantitative trait loci associated with egg quality, egg production, and body weight in an F2 resource population of chickens. Anim Genet 37:106-117.

Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-3.0. 1996-2004 <http://www.repeatmasker.org>.

Tuiskula-Haavisto, M., Honkatukia, M., Vilkki, J., de Koning, D.J., Schulman, N.F., Maki-Tanila, A. 2002. Mapping of quantitative trait loci affecting quality and production traits in egg layers. Poultry Science 8: 919-927

Wang, Y and Leung, F.C.C. An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics 20: 1170-1177.

Wakefield, M.J. 1998. Internet comparative mapping resources. ILAR Journal online 39: 2/3.

Weber, B., Collins, C., Kowbel, D., Reiss, O and Hayden, M.R. 1991. Identification of

multiple CpG islands and associated conserved sequences in a candidate region for the Huntington disease gene. Genomics 11: 1113-1124.

Webster, M.T., Axelsson, E and Hans Ellegren. Strong regional biases in nucleotide substitution in the chicken genome. Mol. Biol. Evol. 23: 1203-1216.

Takai, D and Jones, P.A. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci 99: 3740-3745.

Talbot, C.J., Nicod, A., Cherney, S.S., Fulker, D.W., Collins, A.C and Flint, Jonathan. 1999. High resolution mapping of quantitative trait loci in outbred mice. Nature Genetics 21: 305-308.

Yoder, J.A., Walsh, C.P and T.H. Bestor. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13: 335-340.

Yu, J., Tong, S., Shen, Y and Kao, F-T. 1997. Gene identification and DNA sequence analysis in the GC-poor 20 megabase region of human chromosome 21. Proc. Natl. Acad. Sci 94: 6862-6867.

Website references

http://www.ebi.ac.uk\emboss/cpgplot

http://www.ensembl.org.

http://www.ncbi.nlm.nih.gov

http://www.uniprot.org.

Table 4.1 Total number of repeats in BW QTL region

Type of repeat	Total Number
DNA	371
LINE	2072
Low complexity	2020
LTR	472
Satellite	14
Simple repeat	1637
SINE	0
CpG islands	179

Table 4. 2 Gene density and number of CpG islands in the QTL region for growth and body weight on chicken chromosome 4

Region (Mb)	Number of genes	Number of CpG islands	GC content (%)	Type of repeat (%)						
				SINE	LINE	LTR	DNA	Simple	Low complexity	Total
68-70	36	66	39.64	0	4.10	0.73	0.58	0.48	0.56	6.45
70-72	11	8	35.76	0	7.97	2.03	1.73	0.71	0.80	13.24
72-74	12	25	37.05	0	7.98	2.41	1.89	0.59	0.71	13.58
74-76	23	12	37.60	0	4.37	0.86	1.21	0.50	0.57	7.51
76-78	18	32	39.36	0	4.41	0.78	0.72	0.44	0.54	6.89
78-80	38	36	38.25	0	4.73	0.66	0.66	0.48	0.64	7.17

Table 4. 3 Association between genes and CpG islands in a QTL region for BW on GGA4

Gene	Location	Number of CpG islands	Average Length of islands(bp)	Average % GC content	Average value of O/E
Glucosamine 6-phosphate deaminase 2	68069542-68074986	None	n/a	n/a	n/a
Hypothetical protein	68076993-68093181	1	270	65.93	0.85
YIPF7	68114222-68126201	None	n/a	n/a	n/a
Potassium channel tetramerisation domain	68164838-68249317	2	766	63.71	0.82
CCDC4	68622758-68789343	6	674	69	0.95
ATP8A1	68623269-68712508	1	482	70.12	1.05
Solute carrier family	68807583-68829094	None	n/a	n/a	n/a
Transmembrane protein	68841916-68849305	1	258	69.38	0.81
Phox 2b protein	68910682-68911857	1	625	66.08	1.14
Ubiquitin carboxyl terminal hydrolase	69115200-69119747	2	340	74	0.91
Amyloid beta A4 precursor	69198002-69282152	1	365	64.38	0.87
Nicotinic Acetylcholine receptor	69438067-69442551	None	n/a	n/a	n/a
Rho-related GTP binding protein	69463609-69465218	None	n/a	n/a	n/a
N4BP2_human	69491892-69514419	None	n/a	n/a	n/a
Similar to KIAA0648	69543098-69611626	None	n/a	n/a	n/a
Hypothetical protein	69627827-69643002	None	n/a	n/a	n/a
Replication factor C	69699108-69729077	None	n/a	n/a	n/a
Beta Klotho	69739130-69751382	None	n/a	n/a	n/a
60S ribosomal protein	69753363-69758895	1	308	63.64	0.74
Lipoic acid synthetase	69759164-69770125	2	486	70.67	1.00
UDP-glucose dehydrogenase	69773204-69790271	1	543	67.59	0.95
Phosphoglucomutase	70026641-70034600	1	353	69.69	0.82
Hypothetical protein	70066724-70087773	1	554	74.73	1.03

Centaurin delta	70493178-70499514	None	n/a	n/a	n/a
Ankyrin repeat and pleckstrin protein	70523211-70605649	None	n/a	n/a	n/a
Protocadherin 7 precursor	71884327-72045318	2	520	64.09	0.89
Stromal interaction molecule 2 precursor	73337469-73397344	2	323	66.8	0.89
TBC1 domain family member	73410751-73452645	None	n/a	n/a	n/a
Cholecystokinin type A receptor	73486803-73492556	None	n/a	n/a	n/a
Recombining binding protein suppressor	73507742-73532135	None	n/a	n/a	n/a
Solute carrier family 34	73610600-73627741	None	n/a	n/a	n/a
Hypothetical protein	73685378-73702506	1	464	71.77	1.05
Zinc finger protein 4	73704017-73712663	None	n/a	n/a	n/a
Similar to phosphatidylinosol	73715170-73730521	1	603	75.46	1.08
Similar to SLA/LP autoantigen	73733413-73765695	None	n/a	n/a	n/a
Leucine repeat precursor	73784369-73816045	1	548	82.12	0.95
Extracellular superoxide dismutase	73901427-73902577	None	n/a	n/a	n/a
Putative pre RNA splicing factor	73939199-73993184	1	492	70.93	1.02
Peroxisome proliferative receptor	74334352-74402478	None	n/a	n/a	n/a
Probable G protein coupled receptor	74751846-74812945	2	435	62.70	0.85
Kv channel interacting protein 4	75243899-75291452	None	n/a	n/a	n/a
Slit2 protein	75329564-75432865	None	n/a	n/a	n/a
Slit 2 fragment	75576358-75577593	None	n/a	n/a	n/a

Transcription factor mlr1	76213749-76228730	None	n/a	n/a	n/a
Condensin complex subunit 3	76262479-76296774	None	n/a	n/a	n/a
LIM domain binding 2	76558349-76692868	None	n/a	n/a	n/a
Hypothetical protein	76753717-76791187	1	989	73	1.08
Hypothetical protein	76824043-76872764	None	n/a	n/a	n/a
FGF binding protein 2	76877649-76879959	None	n/a	n/a	n/a
FGF binding protein 1	76893451-76895978	None	n/a	n/a	n/a
ADP ribosyl cyclase 1	76914383-76934076	None	n/a	n/a	n/a
ADP ribosyl cyclase 2	76950097-76963596	None	n/a	n/a	n/a
F-box leucine repeat protein	76967279-77000199	1	812	67.24	0.81
Complement C1q tumor necrosis protein	77056816-77058398	None	n/a	n/a	n/a
Cytoplasmic polyadenylation element binding protein	77144681-77191399	1	206	73.30	1.09
Protein FAM 44A	77693826-77700563	1	793	74.91	0.98
Homeobox protein Nkx 3.2	77739308-77741856	2	684	71.18	1.02
Ras-related protein rab-28	77761569-77822575	1	488	68.65	1.01
Mast cell immunoreceptor signal transducer	78535929-78558995	None	n/a	n/a	n/a
WD repeat protein 1	78620173-78640046	2	279	68.29	0.96
Solute carrier family 2	78685293-78769703	1	773	72.70	0.93
D(1B) dopamine receptor	78804662-78806357	1	259	71.04	0.77
Otoperin 1	78816109-78828645	None	n/a	n/a	n/a
Hypothetical	78832641-78835499	1	202	72.77	0.79
Cell growth regulating nucleolar protein	78847613-78854098	1	290	67.24	1.02

Similar to zinc finger protein	78854122-78871724	1	265	64.15	0.66
LZ1	78889737-78910671	None	n/a	n/a	n/a
Syntaxin 18	78912251-78967335	2	279	70.70	0.89
Homeobox protein ghox-7	79122300-79124789	2	515	65.08	1.06
Cytoxine like protein 1	79195293-79197978	None	n/a	n/a	n/a
Serine/ Threonine kinase 32B	79223802-79371542	1	313	73.48	0.82
Ellis-van Creveld syndrome 2	79402097-79465414	1	247	78.14	0.67
Ellis-van Creveld syndrome	79465505-79517426	1	358	71.79	0.89
Collapsin response mediator protein 1	79523337-79566469	3	467	65.57	0.94
Multiple coiled gabagr1 protein	79680636-79720050	None	n/a	n/a	n/a
Wolframin	79802197-79834712	None	n/a	n/a	n/a
Serine/threonine protein phosphatase	79849236-79972466	1	202	50.99	0.77
Epididymis specific alpha-mannosidase precursor	80024999-80051924	1	200	68.00	0.74
TBC1 domain family member 14	80176777-80238816	4	483	68.71	1.08
MGC21874 protein	80242969-80247177	3	676	69.74	0.94
GrpE-Like 1	80251485-80257891	2	292	62.47	0.82

Table 4.4**.** Classification of selected genes with homologs on GGA4

Gene	Gene Ontology	Position on GGA4	Location of closest CpG island in GGA4	Tissue Specificity
Rhotekin	Molecular Function GTP binding Nucleotide binding Rho GTPase binding Cellular Component Intracellular Biological Process Apoptosis Signal transduction	69430821-69434843	69410837-69411116	Widespread
Retinoblastoma binding protein 5	Molecular Function Protein binding Cellular Component Histone methyltransferase Nucleus	69833079-69837108	69834068-69834289	Widespread

Tim44 mitochondrial precursor	Cellular Component Mitochondrial matrix Presequence translocase-associated import motor Molecular Function Protein transporter activity Biological Process Protein import into mitochondrial matrix	72758679-72758702	72829042-72829275	Unclear
G protein receptor 125	Molecular Function Transferase activity Protein binding Receptor activity Cellular Component Integral to membrane Biological Process Cell adhesion Signal transduction Neuropeptide signaling	74751846-74812945	72829042-72829275	Unclear

CG4080-PA	Molecular Function Protein binding Zinc ion binding Metal ion binding Cellular Component Nucleus Biological Process Ubiquitin	78345008-78345033	78349998-78350431	Expressed in the male germline germline (tissue specific)
Forkhead box I2	Biological Process Transcription Transcription antitermination Regulation of transcription, DNA dependent Molecular Function DNA binding	79223168-79223188	79223085-79223257	Widespread

	transcription factor activity
Cellular Component	
	Nucleus

CHAPTER 7

GENERAL CONCLUSIONS

Traits of economic importance are controlled by several genes and the environment. The detection of quantitative trait loci (QTL) and the genes underlying them are therefore important for the improvement of these traits. A study was conducted to locate putative candidate genes on GGAZ by orthologous comparison of QTL regions on GGAZ with the mouse and human genomes. Primer sequences from markers flanking QTL regions were blasted against the chicken genome using BLASTN (http://www.ensembl.org). Forty six chicken genes together with 91 mouse and 60 human genes were identified in this study. The annexin A1 gene, follistatin and neuronal acetylcholine receptor gene (nAChR) were some of the genes identified in this work. The nAChR gene is located at a QTL region for abdominal fat and could be used a therapeutic agent for feed intake and obesity. A second study was conducted to detect QTL regions for growth and skeletal traits in an F_2 population selected for high and low growth. QTL for age-related body weight (BW), shank length (SL) and shank diameter (SD) were localized in 695 individuals. A pleiotropic QTL on GGA4 explained 7-11% of the variance and affected BW at 5 to 9 weeks, and SL and SD at 9 weeks. A male –specific BW QTL was detected on GGA3 at 173 cM. The QTL on GGA4 had the strongest on SL and SD, and explained 18% and 21%, respectively of the variance. A third study located QTL for carcass composition and fat on GGA 2, 3, 4,5,10 and 26. Differentially expressed genes in the QTL regions included SOD and fat-1. The Spot 14 gene is

associated with abdominal fat in chickens. The fourth study was conducted to characterize a major QTL region on GGA4 and to identify candidate genes in this region by CpG island detection and comparative mapping. One hundred and nine known genes and 179 CpG islands were located at this locus. Six putative novel genes were identified by blasting genes from 23 orthologous species against the chicken genome. A putative ortholog of the rhotekin gene was detected. Rhotekin is a housekeeping gene with acetylcholine activity and may be linked with nAChR.

REFERENCES

Abasht, B., Pittel, F., Lagarrigue, S., Le Bihan-Duval, E., Le Roy, P., Demeure, O., Vignoles, F., Simon, J.M., Cogburn, L., Aggrey, S., Vignal, A., Douaire, M. 2006. Fatness QTL on chicken chromosome 5 and interaction with sex. Genetics, Selection Evolution 38: 297-311.

Aitman, T.J., Glazier, A.M., Wallace, C.A., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., Al-Majali, K.M., Trembling, P.M., Mann, C.J., Shoulders, C.C., Graf, D., St. Lezin, E., Kurtz, T.W., Kren, V., Pravenec, M., Ibrahimi, A., Abumrad, N.A., Stanton, L.W and Scott, J. 1999. Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive. Nature genetics 21: 76- 83.

Ankra-Badu, GA and Aggrey, SE. 2005. Identification of candidate genes at quantitative trait loci on chicken chromosome Z using orthologous comparison of chicken, mouse and human genomes. In Silico Biol. 5: 593-604.

Antequera, B and Bird, A. 1993. Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci 90: 11995-11999.

Ashburner, M., et al.; The Gene Ontology Consortium (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25-29.

Ashikawa, I. 2002. Gene-associated CpG islands and the expression pattern of genes in rice. DNA Research 9: 131-134.

Bellamy, R., Ruwende, C., Corrah, T., McAdam, K.P.W.J., Whittle, H.C and Hill A.V.S. 1998. Variations in the Nramp1 and susceptibility to tuberculosis in West Africans. New England Journal of Medicine 338: 640-644.

Bennet, A.K., Hester, P.Y and Spurlock, D.E.M. 2006. Polymorphisms in vitamin D receptor, osteopontin, insulin-like growth factor 1 and insulin, and their associations with bone, egg and growth in a layer-broiler cross in chickens. Animal Genetics 37: 283-286

Borevitz, J.O and Chory, J. 2004. Genomic tools for QTL analysis and gene discovery. Current opinion in plant biology 7: 132-136

Broman, K.W. (2001). Review of statistical methods for QTL mapping in experimental crosses. Lab Animal 30: 44-52.

Bull, S.B., John, S. and Briollais, L. 2005. Fine mapping by linkage and association in nuclear family and case-control designs. Genetic epidemiology 29: S48-S58.

Cross S.H., Clark, V.H and Bird, A.P. 1999. Isolation of CpG islands from large genomic clones. Nucleic Acids Research 27: 2099-2107.

Darvasi, A. (1998). Experimental strategies for the genetic dissection of complex traits in animal models. Nature genetics 18: 19-24.

Darvasi, A. and Soller, M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141: 1199-1207.

Davignon J., R.E. Gregg, C.F. Sing. 1988. Apolipoprotein E polymorphism and atherosclerosis. Arteriosclerosis 8:1-21.

De Koning D.J., Janss, L.L.G., Rattink, A.P., Van Oers, P.A.M., De Vries, B.J., Groenen, M.A.M., Van DE Poel, J.J., de Groot, P.N., Brascamp, E.W., Van Arendonk, J.AM. 1999. Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (Sus scrofa). Genetics 152: 1679-1690.

De Vienne, D and Causse, M. 2003. Mapping and characterising quantitative trait loci. pp 89-124. Molecular markers in plant genetics and biotechnology. Dominique de Vienne (ed). Enfield, NH: Science Publishers

Doerge, R.W., Zeng, Z-B and Weir, B.S. 1997. Statistical issues in the search for genes affecting quantitative traits in experimental populations. Statistical Science 12: 195-219.

Dunn, I.C., Miao, Y.W., Morris, A., Romanov, M.N., Wilson, P.W and Waddington, D. 2004. A study of association between genetic markers in candidate genes and reproductive traits in one generation of a commercial broiler breeder hen population. Heredity 92: 128-134.

Dunnington, E.A., Gal, O., Plotsky, Y., Haberfeld, A., Kirk, T., Goldberg, A., lavi, U.,Cahaner, A., Siegel, P.B and Hillel, J. 1990. DNA fingerprints of chickens selected for a high and low body-weight for 31 generations. Genetics 21: 247-257.

Falconer, D.S and Mackay, T.F.C. 1996. Introduction to quantitative genetics 4th edition Essex, England, Longman pp 356-376.

Gillespie, G.A.J., Somlo, S., Germino, G.G., Weinstat-Saslow, D and Reeders, S.T. 1991. CpG islands in the region of an autosomal dominant polycystic kidney disease locus defines the 5' end of a gene encoding a putative proton. Proc Natl Acad Sci 88: 4289-4293.

Gunnarson, U., Kerje, S., Jensen, P and Andersson, L. 2006. Fine mapping of a large growth QTL in an intercross between the red junglefowl and white leghorn chickens. Plant & Animal genomes XIV conference. http:/www.intl-pag.org/14/abstracts/PAG14_P553.html.

Haseman, J.K and Elston, R.C. 1972. The investigation of linkage between a quantitative trait and a marker lous. Behav. Genet 2: 3-19.

Hoeschele, I and Vandraden, P.M. 1993. Bayesian-Analysis of linkage between genetic-markers and quantitative trait loci. 2. Combining prior knowledge with experimental-evidence. Theoretical and applied genetics 85: 946-952

Ikeobi, C.O.N., Wooliams, J.A., Morrice, D.R., Law, A., Windsor, D., Burt, D.W and Hocking, P.M. 2002. Quantitative trait loci affecting fatness in the chicken. Animal Genetics 33: 428-435.

Jayakar, S.D. 1970. On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. Biometrics 26: 451-464.

Jennen, D.G.J., Vereijken, A.L.J., Bovenhius, H., Crooijmans, R.P.M.A., Veenendaal, van der Poel, J.J and Groenen, M.AM. (2004). Detection and localization of quantitative loci affecting fatness in broilers. Poultry Science 83: 295-301.

Johansson, M., Ellegren, H and Andersson, L. 1995. Comparative mapping reveals extensive linkage conservation- but with gene order arrangements- between the pig and the human genomes. Genomics 25: 682-690.

Karp, C.L., Grupe, A., Schadt, E., Ewart, S.L., Keane-Moore, M., Cuomo, P.J., Kohl, J., Wahl, L., Kuperman,D, Germer, S., Aud, D., Peltz, G and Wills-Karp. 2000. Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. Nature Immunology 1: 221-226.

Larsen, F., Gundersen, G., Lopez, R and Prydz, H. 1992. CpG islands as gene markers in the human genome. Genomics 13: 1095-1107.

Lee, K-T., Park, E-W., Moon, S., Park, H-S., Kim, H-Y., Jang, G-W., Choi, B-H., Chung, H.Y., Lee J-W., Cheong, I-C., Oh, s-j., Kim, H., Suh, D-S and T-H Kim. 2006. Genomic sequence analysis of a potential QTL region for fat trait pig chromosome 6. Genomics 7: 218-224.

Liu, B.H. 1998. Statistical genomics: linkage, mapping, and QTL analysis. Boca Raton: CRC Press.

Liu, H-C., Cheng, H.H., Tirunagaru, V., Sofer, L and Burnside, J. 2001. A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. Animal Genetics 32: 351-359.

Korstanje, R and Paigen, B. 2002. From QTL to gene: the harvest begins. Nature genetics 31: 235-236.

Lander, E.S. and Schork, N.J. 1994. Genetic Dissection of Complex Traits. Science 265: 2037-2048

Lynch, M. and Walsh, B. 1998. Genetics and Analysis of quantitative traits. Sinauer associates, inc. Sunderland, Massachussets.

Maestini, E., Tamanini, F., Kioschis, P., Gimbo, E., Marinellini, P., Tibioli, C., D'Urso, M., Palmieri G, Poustka, A and Toniolo, D. An archipelago of CpG islands in Xp28: identification and fine mapping of 20 new CpG islands of the human X chromosome. Hum.Mol.Genet. 4: 275-280.

Maliepaard, C., Sillanpia, M.J., van Ooijen, J.W., Jansen, R.C and Arjas, 2001. Bayesian versus frequentist analysis of multiple quantitative trait loci with an application to an outbred apple cross. Theor Appl Genet 103: 1243- 1253.

McPeek, M.S. 2000. From mouse to human: fine mapping of quantitative trait loci in a model organism. Proceedings of the national academy of sciences. 97: 12389-12390.

Nanda, I., Haaf, T., Schartl, M., Schmid, M and Burt, D.W. 2002. Comparative mapping of Z-orthologous genes in vertebrates: implications for the evolution of avian sex chromosomes. Cytogenet Genome Res 99: 178-184.

Neimann-Soressen, A. and Robertson, A. 1961. The association between blood groups and several production characters in three Danish cattle breeds. Acta. Agr. Scand. 11: 163-196.

Nelson, M.R., Kardia, S.L.R., Ferrell, R.E and Sing, C.F. 1999. Influence of apolipoprotein E genotype variation on the means, variances and correlations of plasma lipids and apoliproteins in children. Annals of Human Genetics 63: 311-328

Nguyen, T.T.T., Klueva, N., Chamareck, V., Aarti, A., Magpantay, G., Millena, A.C.M., Pathan, M.S and Nguyen, H.T. (2004). Saturation mapping of QTL regions and identification pf putative candidate genes for drought tolerance in rice. Mol. Gen. Genomics 272: 35-46.

Pflieger, S., Lefebvre, V and Causse, M. 2001. The candidate gene approach in plant genetics: a review. Molecular Breeding 7: 275-291.

Plotsky, Y., Cahaner, A., Haberfeld, A., Lavi, U and Hillel, J. 1990. Analysis of genetic association between DNA fingerprint bands and quantitative traits using DNA mixes. Proc 4th world congress- on genetics applied to Livestock Prod 13: 133-136. Edinburgh.

Remington, D and Puruggan, M.D. 2003. Candidate genes, quantitative trait loci, and functional trait evolution in birds. Int. J. Plant Sci 164: S7-S20.

Sasaki O, Odawara S, Takahasi H, Nisawara K, Oyamada Y, Yamamoto R, Ishii K, Nagamine Y, Takeda H, Kobayashi E, Furukawa T. 2004. Genetic mapping of quantitative trait loci affecting body weight, egg character and egg production in F2 intercross chickens. Anim Genet 35: 188-194.

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8: 552-560.

Schreiweis, M.A., Hester, P.Y., Settar, P and Moody, D.E. 2006 Identification of quantitative trait loci associated with egg quality, egg production, and body weight in an F2 resource population of chickens. Anim Genet 37:106-117.

Serebovsky, A.S. and Petrov, S.G. 1930. On the composition of the plan of the chromosomes of the domestic hen. Zhurnal experimental'noy biologii 6: 157-180.

Sewalem, A., Morrice, D.M., Law, A., Windsor, D., Haley, C.S., Ikeobi, C.O.N., Burt, D.W and Hocking, P.M. 2002. Mapping of quantitative trait loci for body weight at three, six and nine weeks of age in a broiler layer cross. Poultry Science 81: 1775-1781.

Stoll, M., Kwitek-Black, A.E., Cowley, A.W., Harris, E.L., Harrap, S.B., Krieger, J.E., Printz, M.P., Provoost, A.P., Sassard, J and Jacob, H.J. (2000). New target regions for human hypertension via comparative genomics. Genome Res. 10: 473-482.

Stratil, A and Geldermann, H. 2004. Analysis of porcine candidate genes from selected QTL regions affecting production traits. Animal Science Papers and reports 22: 123-125.

Suchyta, S.P., Cheng, H.H., Burnside, J and Dodgson, J.B. 2001. Comparative mapping of chicken anchor loci orthologous to genes on human chromosomes 1, 4 and 9. Animal Genetics 32: 12-18.

Tabakoff, B., Bhave, S.V and Hoffmann, P.L. 2003. Selective breeding, quantitative trait locus analysis, and gene arrays identify candidate genes for complex drug-related behaviors. The Journal of Neuroscience 23: 4491-4498.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E.S. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. 28: 286-289.

Tuiskula-Haavisto, M., Honkatukia, M., Vilkki, J., de Koning, D.J., Schulman, N.F., Maki-Tanila, A. 2002. Mapping of quantitative trait loci affecting quality and production traits in egg layers. Poultry Science 8: 919-927.

Uimari, P and Hoeschele, I. 1997. Mapping linked quantitative trait loci using Bayesian analysis and Markov Chain Monte Carlo Algorithms. Genetics 146: 735-743.

Vallejo, R.L., Bacon, L.D., Liu, H.C., Witter, R.L., Groenen, M.A.M., Hillel, J., Cheng, H.H. 1998. Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F-2 intercross chickens. Genetics 148: 349-360.

Van Kaam, J.B.C.H.M., Groenen, M.A.M., Bovenhius, H., Veenendaal, A., Vereijken, A.L.J and Van Arendonk, J.A.M. 1999. Whole genome scan in chickens for

quantitative trait loci affecting growth and feed efficiency. Poultry Science 78: 15-23.

Wayne, M.L and McIntyre, L.M. Combining mapping and arraying: an approach to candidate gene identification. Proceedings of the national academy of sciences 12: 14903-14906.

Weber, B., Collins, C., Kowbel, D., Reiss, O and Hayden, M.R. 1991. Identification of multiple CpG islands and associated conserved sequences in a candidate region for the Huntington disease gene. Genomics 11: 1113-1124.

Weller, J.I. 2000. Quantitative trait loci analysis in animals. CABI publishers, Wallingford, UK.

Weterman, M.A.J., Wilbrink, M and van Kessel, A. 1996. Fusion of the transcription factor TFE3 gene to anovel gene, PRCC, in t(X;1)(p11;q21)-positive papillary renal call carcinomas. Proc Natl Acad Sci 93: 15294-15298.

Yoshizawa, K., Inaba, K., Mannen, H., Kikuchi, T., Mizutani, M and Tsuji, S. 2002. Fine mapping of the muscular dystrophy (AM) gene on chicken chromosome 2q. Animal Genetics 35: 397-400.

Yu, J., Tomg, S., Shen, Y and Kao, F-T. 1997. Gene identification and DNA sequence analysis in the GC-poor 20 megabase region of human chromosome 21. Proc. Natl. Acad. Sci. 94: 6862-6867

Zhu, J.J., Lillehoj, H.S., Allen, P.C., Van Tassell, C.P., Sonstegard, T.S., Cheng, H.H., Pollock, D., Sadjadi, M., Min, W and Emara. 2003. Mapping and quantitative trait loci associated with resistance to coccidiosis and growth. Poultry Science 82: 9-16.

Website references:

http://www.informatics.jax.org/silver/chapters

Appendix A. Comparative mapping of gene-poor CpG island regions on GGA4 with other species

GGA4 Location (cM)	Comparative Species	Gene Location	Genes
68-70	D. novemcinctus	Gene Scaffold_1361: 1241-5261	UDP transporter solute carrier family 35 family
	L. africana	Gene Scaffold_4707:45489-49503	RNA binding with multiple splicing RBP MS heart RRM expressed sequences hermes family
	D. melanogaster	3L:17714289-17716308	Adenosine deaminase-related growth factor B CG5998-PA
	T. rubripes	Scaffold_124: 706692-710710	Rhotekin (Homolog of HS Rhotekin-2)
	T. nigroviridis	1: 4482901-4866922	Mediator of RNA polymerase II transcription subunit 12 thyroid hormone receptor
	M. mulatta	7: 97709920-97713953	BAZ1A_human (Bromodomain adjacent to zinc finger domain 1a ATP utilizing chromatin assembly and factor 1)

	Mus musculus	1: 134258930-134260958	Retinoblastoma binding protein 5
70-72	C. intestinalis	12p: 1430441-1434460	Zinc finger
	D. melanogaster	X: 20581183-20583209	Innexin shaking B
	D. melanogaster	X: 17998691-18000720	Frequenin-1
	D. melanogaster	2L: 305913-307931	1, phosphatidylinositol 4,5 biphosphate phosphodiesterase classes I and II
72-74	G. aculeatus	IX: 1208984-1211035	Protocadherin precursor family
	H. sapiens	4: 30331199-30333699	Protocadherin 7 Precursor
	M. mulatta	5: 25612725-25615225	Protocadherin
	M. domestica	5: 195434566-195781954	Protocadherin 7 Precursor
	M. musculus	5: 58007062-58009615	Protocadherin 7 Precursor
	P. troglodytes	4: 30982552-30985052	Protocadherin 7 Precursor

	C. savignyi	Reftig 20: 696671-700695	TIM44 mitochondrial precursor
	X. tropicalis	Scaffold_85: 108175-112200	Noelin3-precursor (optimedin)
	C. familiaris	1: 121306668-121308691	cDNA product: hypothetical RUVA domain 2 structure containing protein
	C. elegans	II: 1277979-1280000	serpentine receptor class Z family member
	X. tropicalis	Scaffold_925: 67674-86095	Cicardian pacemaker protein
	D. melanogaster	2L: 6943371-6945393	Circulating catholic antigen CCA protein
	P. troglodytes	19: 52702135-52704153	Myosin heavy chain
	D. melanogaster	X: 19575485-19579511	Ribonuclease P subunit family
	H. sapiens	5: 132177020-132179041	Ankyrin repeat domain containing protein 43
	M. mulatta	6: 129220827-12922848	Ankyrin repeat domain containing protein 43

	M. domestica	2: 515166262-515168290	Carboxylase precursor family
	R. norvegicus	3: 169519469-169521495	Solute carrier organic anion transporter family member 4A1
	T. nigroviridis	11: 827548-831569	RNA polymerase II elongative factor eleven nineteen lysine rich leukemia family
74-76	C. elegans	V: 16043938-16045955	Histone H3 and H4
	D. novemcinctus	Genescaffold_2711: 52257-55267	Plaur domain containing 3 precursor family
	H. sapiens	17: 4034875-4036896	YRNA
	O. cuniculus	Genescaffold_2799: 43159-47174	Cyclin dependent kinase inhibitor family
	P. troglodytes	17: 4338207-4340228	YRNA
	C. intestinalis	9q: 631039-633061	ORM1 family
	M. mulatta	5: 17166494-17168546	Probable G coupled receptor

			124 precursor tumor endothelial marker 5 family
	M. musculus	5: 50247812-50250020	G protein coupled receptor 125
	T. rubripes	Scaffold_85: 78703-82722	Homolog of Homo Sapiens Malonyl-CoA decarboxylase mitochondrial precursor
	D. novemcinctus	Gene scaffold_2799: 7748-11763	Insulin gene enhancer islet family
	L. africana	Gene scaffold_6143: 237056-241071	Phosphatase methylesterase family
	M. mulatta	9: 96903980-96906001	Proto-oncogene FRAT 1 (frequently rearranged in advanced T-cell lymphomas)
	M. musculus	3: 89294389-89296418	Metaxin 1
	P. troglodytes	10: 100803061-100805082	Proto-oncogene FRAT 1 frequently rearranged in advanced T-cell lymphomas
	R. norvegicus	7: 88345479-88347500	A5D3 protein

76-78	H. sapiens	20: 51986058-51988081	Breast carcinoma amplified sequence 1
	B. taurus	29: 43372070-43376116	Potassium voltage gated channel subfamily KQT member 1
	C. elegans	I: 2296408-2298424	Tpx2 related alternative variant family
	C. savignyi	Reftig_11: 883127-887144	Ras interacting 1 rain family
	D. novemcinctus	Gene scaffold_2799: 1-2549	MPV17 family
	R. norvegicus	3: 61553608-61555630	Similar to ubiquitin conjugating enzyme UbcM2
	X. tropicalis	Scaffold_85: 1048809-1052830	Pre-B-cell leukemia transcription factor 1
	C. intestinalis	1q: 2422797-2424812	Cell division cycle 5
	O. cuniculus	Gene scaffold_2799: 162984-167005	Chemokine receptor type 5 antigen
	T. rubripes	Scaffold_46: 658972-662995	Homolog of Homo Sapiens “splice isoform 1 of polypeptide N-acetylgalactosaminyl transferase 13”

	B. taurus	10: 38711343-38713367	Phosphatidylinositol glycan class B
	M. musculus	X: 16288124-16290153	EF- hand domain containing 2
	B. taurus	13: 25707779-25711851	Homeobox protein Nkx-2
	C. savignyi	Reftig_48: 401353-405369	CCDC11
	D. rerio	14: 1303251-1305447	Bagpipe homeobox homolog 1
	D. melanogaster	3R: 1339237-13394429	H6-like homeobox CG5832-PA
	G. aculeatus	IV: 954517-956710	Homeobox protein Nkx-2
	M. mulatta	5: 8353301-8355322	Homeobox protein Nkx-2
	M. musculus	17: 26566066-26568139	Nk2 transcription factor related
	P. troglotydes	4: 13441823-13443844	Homeobox protein Nkx-2 family
	R. norvegicus	10: 16607242-16609279	Homeobox protein Nkx-2 family

78-80	C. familiaris	3: 72479685-72481704	WD repeat 1 actin interacting 1 family
	D. melanogaster	3R: 22361205-22363221	Protein lap 4
	C. elegans	V: 1609937-1611963	Nuclear hormone receptor family member NHR family
	C. savignyi	Reftig_155: 464023-468042	NOT5 family
	D. melanogaster	X: 10554558-10558578	MIDN family
	G. aculeatus	XVII: 7684360-7891914	Rac GTPase activating 1 MGCRACGAP Ellis Van Creveld syndrome LIMBIN family Homeobox protein family Syntaxin 18
	M. domestica	5: 221673015-221809636	Homeobox protein
	T. rubripes	Scaffold_37: 1096048-1100073	Homolog of Homo Sapiens EF hand calcium binding protein 1
	C. intestinalis	10p: 1602928-1604960	Histone H2A family

	D. melanogaster	3L: 9360491-9362517	CG4080-PA (Homologous to cellular modulator of immune recognition
	T. rubripes	Scaffold_430920-434940	Homolog of Homo Sapien protein CGI-117
	C. familiaris	5: 35188553-35190663	Similar to S-phase 2 protein
	H. sapiens	19: 55711888-55713915	Leucine rich repeat containing protein 4B precursor
	P. troglodytes	19: 53019305-53021332	Leucine rich repeat containing protein 4B precursor
	H. sapiens	10: 129424420-129426440	Forkhead box I2
	M. mulatta	7: 40459028-40461055	Nuclear localized factor 1 homolog
	M. musculus	11: 55046665-55048691	Solute carrier family 36
	G. aculeatus	IX: 6432063-6434080	Potassium voltage gated channel subfamily member 1
	M. musculus	7: 46263877-46265896	Potassium voltage gated channel shaw related subfamily member 1

	O. cuniculus	Genescaffold_2799: 1-2750	Serine / threonine kinase family
	R. norvegicus	1: 96927593-96929612	Potassium voltage gated channel subfamily member 1
	T. nigroviridis	21: 5052559-5054576	Potassium voltage gated channel subfamily member 1
	L. africana	Genescaffold_2799: 37113-41132	Spermatogenesis genesis associated 7
	M. musculus	14: 45803670-45805696	Sterile alpha motif domain containing 4