NOTE TO USERS

PHOSPHOENOLPYRUVATE CARBOXYKINASE AND ORNITHINE DECARBOXYLASE GENES: ALLELIC VARIATIONS AND ASSOCIATIONS WITH TRAITS IN POULTRY

by

Reza Parsanejad

Department of Animal Science
Macdonald Campus of McGill University
Montreal, Canada

November, 2003

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Your file Votre référence
ISBN: 0-612-98344-7
Our file Notre référence
ISBN: 0-612-98344-7

Suggested Short Title:

CANDIDATE GENES AND TRAIT ASSOCIATIONS

ACKNOWLEDGMENTS

I would like to express my sincere thanks and gratitude to Dr. Urs Kuhnlein, my thesis supervisor, for his keen intellect, inspiration, encouragement, valuable guidance and professional advice throughout the tenure of my doctoral program. I particularly appreciate his friendly leadership and respect throughout the course of this study. I was privileged to have been under his tutelage.

Sincere thanks are also extended to Dr. David Zadworny, whose advice was instrumental in bringing this thesis to completion. His door was always open for questions. I am also deeply grateful to the other distinguished members of my dissertation committee: Dr. Xin Zhao (Chair of Animal Science Department) and Dr. John Flannan Hayes who contributed their time, knowledge and expertise.

I express my gratefulness to Ms. Barbara Stewart and Sandra Nagy, Animal Science Department secretaries, for all the administrative and secretarial help they extended to me. I would like to thank Ms. Dana Praslickova, Ms. Katja Linher, Ms Jovette Bochard, Ms. Twinkle Masilamani, Mr. Babu Golapodi, Dr. Gregoy Bedecarrats, Dr. Sammy Aggrey, Mr. Nabil Fetni and Mr. Benoit Leclerc for freely sharing their knowledge and experience regarding the technical aspects and data analysis, and also providing me the great emotional support and friendly environment. The cooperation and assistance given by Ms. Marilyn Richard who was always ready to help and translated the abstract into French are gratefully acknowledged.

A heartfelt gratefulness is extended to my beloved mother, Omehlila (Azar) Zaroori, for all her love and morals she instilled in me, and for her constant assurance that

"hard work breeds success". Thank you so very much for believing in me. Your sacrifices made all of this possible. You are a great mother. An extraordinary thanks goes out to my sisters, Farah and Faranak and my brother, Hasan, whose optimistic outlook and support always gave me the confidence boost I needed.

This thesis is dedicated to my wonderful, lovely wife for her understanding, encouragement and help whenever I needed the most. Her encouraging words were a blessing. It is also dedicated to my lovely daughters, Helia and Nyousha, for their patience and understanding without which this study would not have been successfully completed. I wish they forgive me for all the time that I missed to be with them.

ABSRACT

The objectives of this study were to identify genetic variants, develop the respective haplotypes (combination of alleles) and investigate the association of identified variants with economically important traits in two candidate genes. The first gene was Phosphoenolpyruvate carboxykinase (PEPCK) which is a key regulatory enzyme of gluconeogenesis. The second candidate gene, Ornithine decarboxylase (ODC), is a rate-limiting enzyme in polyamine biosynthesis. It has a significant role in DNA synthesis and cell proliferation. We first analyzed the genetic variability of PEPCK-C, the gene which codes for the cytosolic form of PEPCK. A 3792 bp segment of 5'-region of the PEPCK-C gene (pos. -1723 to 2069) was sequenced in 32 White Leghorn chickens (a total of 64 genomes). A total of 19 single nucleotide polymorphisms (SNPs) were identified. We then analyzed the genetic variability of ODC. A 5 kb sequence of 3' end of the gene was sequenced in 20 White Leghorn chickens (a total of 40 genomes). A total of 63 variant sites were identified. The rate of insertion/deletion in ODC was 16%, whereas neither deletions nor insertions were present in PEPCK-C. Gene trees were constructed for both genes assuming maximal parsimony. This led to the delineation of 6 haplotypes in PEPCK-C. Two of the SNPs coincided with RFLP detectable by the restriction enzymes *Aci*I and *Bst*EII, respectively. Three haplotypes in ODC were defined. In the next step, White Leghorn chickens from a non-selected closed population were typed for these two PEPCK-C RFLP. The two RFLP gave rise to three alleles (or haplotype classes), which in turn defined six genotypes. A comparison of genotypes revealed significant differences in feed efficiency (FE) and residual feed consumption (RFC). There was significant

interaction between PEPCK-C genotypes and mitochondrial PEPCK (PEPCK-M) genotypes defined by an RFLP. The latter enzyme catalyzes the same reaction, but is located in the matrix of the mitochondria and is encoded by a different nuclear gene. Interaction was evident from an analysis of the egg weight and egg-specific gravity in the early phase of egg laying. In addition, significant genetic disequilibria were observed for two of the three alleles of the cytosolic PEPCK and between one of these alleles and the mitochondrial PEPCK RFLP. This indicates that variations of genes in the gluconeogenesis pathway may affect feed utilization and egg production traits, as well as reproductive fitness. In addition, the two RFLP in the ODC gene gave rise to three alleles (or haplotype classes), which in turn defined six genotypes. A comparison of genotypes revealed significant differences in the body weight, the age of first egg, egg-specific gravity and residual feed consumption.

RESUME

Les objectifs de ce projet étaient d'identifier des variantes génétiques, d'établir les haplotypes (une combinaison d'allèles) et d'investiguer les associations des variantes avec des traits d'importance économique dans deux gènes candidats. Le premier gène était la phosphoenolpyruvate carboxykinase (PEPCK) qui est une protéine clef dans la régulation du métabolisme de la gluconéogénèse. Le deuxième gène candidat, l'ornithine decarboxylase (ODC), est une enzyme limitant la synthèse des polyamines. ODC joue un rôle significatif dans la synthèse de l'ADN et la prolifération des cellules. Nous avons d'abord analysé la variation génétique de PEPCK-C, le gène codant pour la forme cytosolique de PEPCK. Un segment de 3792 pb appartenant à la région 5' du gène PEPCK-C (position –1723 à 2069) a été séquencé chez 32 poulets de type White Leghorn. Un total de 19 Single Nucleotide Polymorphisms (SNPs) ont été identifiés. Nous avons ensuite analysé la variabilité génétique du gène ODC. Un segment de 5 Kb localisé dans la région 3' du gène a été séquencé chez 20 poulets de type White Leghorn. Un total de 63 polymorphismes ont été identifiés. Le taux d'insertion/délétion était de 22% alors qu'aucune insertion/délétions n'ont été localisées dans le gène PEPCK-C. Des arbres génétiques ont été construits pour chaque gène en assumant une parcimonie maximale qui ont conduit à 6 haplotypes pour PEPCK. Deux des SNPs étaient détectables à l'aide des enzymes de restriction *Aci*I et *Bst*EII, respectivement. Trois haplotypes ont été définis pour ODC. A l'étape suivante, le génotype de poulets de type White Leghorn provenant d'une population fermée où aucune sélection n'a été effectuée a été déterminé pour les 2 marqueurs RFLP du gène ODC. Les deux marqueurs RFLP ont résulté en 3

allèles (ou 3 classes d'haplotypes). Une comparaison des génotypes a révélé des différences significatives en efficacité alimentaire {feed efficiency (FE)} et la consommation d'aliments résiduelle {residual feed comsumption (RFC)}. Nous avons noté une interaction significative entre les génotypes des gènes PEPCK-C et PEPCK-M pour les marqueurs RFLP. Le second enzyme catalyse la même réaction mais est localisé dans la mitochondrie et est codé par un gène différent. L'interaction était évidente à partir d'une analyse du poids de l'oeuf et de la gravité spécifique de l'oeuf dans les premières phases de ponte. De plus, des déséquilibres génétiques ont été observés pour 2 des 3 allèles de la PEPCK cytosolique et entre une de ces allèles et le RFLP de la PEPCK des mitochondries. Cela indique que la variation des gènes impliqués dans le processus de gluconéogénèse pourrait affecter l'efficacité alimentaire et les traits reliés à la ponte ainsi que l'efficacité reproductrice. De plus, le génotype de poulets de type White Leghorn provenant d'une population fermée où aucune sélection n'a été effectuée a été déterminé pour les 2 marqueurs RFLP du gène ODC.

CONTRIBUTIONS TO ORIGINAL KNOWLEDGE

Chapter 3, is the first study to document the genetic variations in the PEPCK-C gene in White Leghorn chickens. I identified 19 variants in 4 kb of the 5' end of the gene, delineated haplotypes and established a gene tree. I also localized two restriction sites enzymes, *Aci*I and *BstE*II, for rapid genotyping.

In chapter 4, I provided evidence for an association between PEPCK-C genotypes and residual feed consumption and feed efficiency. I also demonstrated that genotypes were at Hardy-Weinberg disequilibrium. I, then, showed epistasis between the two isozymes PEPCK-C and PEPCK-M.

In chapter 5, 63 variants in 5 kb of the 3' end of the ODC gene were identified in White Leghorn chickens. I, then, defined three haplotypes. A phylogenetic tree was established. Two diagnostic restriction enzymes, *Hind*III and *Msp*I, were used for rapid genotyping and the association between ODC genotypes and production traits was analyzed. It revealed that the ODC gene was associated with body weight, the age of first egg, egg-specific gravity and residual feed consumption.

CONTRIBUTION OF AUTHORS

Part of the thesis has been presented at a conference and has been accepted for publication or submitted for publication. The authors have been involved in these and their contributions to the various articles are as follows:

- R. Parsanejad is the Ph.D. candidate who planned and conducted the experiments, gathered and analyzed the results, and wrote the manuscripts for scientific publication.

- Dr. U. Kuhnlein, the thesis supervisor, provided funding for the research, supervised the laboratory work, statistical analyses, and the preparation of the manuscripts and the thesis.

- Dr. D. Zadworny provided assistance in the correction of the manuscripts.

- Dr. A Tourkamanzehi identified the *Acc*I RFLP and determined the genotypes of PEPCK-M (chapter 4).

LIST OF PUBLICATIONS RELATED TO THE THESIS

Cytosolic phosphoenolpyruvate carboxykinase gene in White Leghorn chickens.

Parsanejad R., D. Zadworny, and U. Kuhnlein.

2000. XXI World's Poultry Congress.

Genetic variability of the cytosolic phosphoenolpyruvate carboxykinase gene in White Leghorn chickens.

Parsanejad R., D. Zadworny, and U. Kuhnlein, 2002.

Poultry Science. 2002. 81(11):1668-70.

Alleles of cytosolic PEPCK: Trait association and interaction with mitochondrial PEPCK in a strain of White Leghorn chickens.

Parsanejad R., A. Torkamanzehi, D. Zadworny and U. Kuhnlein

Poultry Science. 2003. 82(11):1708-15.

Ornithine decarboxylase: Haplotype structure and trait association in White Leghorn chickens.

Parsanejad R., D. Zadworny, and U. Kuhnlein

Poultry Science. 2004. Accepted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION KEY

A	Adenine
AFE	Age of the onset-of laying
ANOVA	Analysis of variance
C	Cytosine
cDNA	complementary deoxyribonucleic acid
d	Day
DNA	Deoxyribonucleic acid
EGM	Egg mass
EWT	Egg weight
FC	Feed consumption
g	Gram
G	Guanine
G6PD	Glucose-6-phosphate dehydrogenase
GDP	Guanosine diphosphate
GH	Growth hormone
GHR	Growth hormone receptor
GPT	Glutamate-pyruvate transaminase
GTP	Guanosine triphosphate
HBWT	Housing body weight
HDR	Rate of egg-laying
IGF-I	Insulin-like growth factor I
kb	Kilo base

KG	Kilogram
LD	Linkage disequilibrium
LDH	Lactate dehydrogenase
MBWT	Mature body weight
MD	Marek's disease
N	Number
ODC	Ornithine decarboxylase
PCR	Polymerase chain reaction
PEP	Phosphoenopyruvate
PEPCK	Phosphoenolpyruvate carboxykinase
PEPCK-C	Cytosolic form of Phosphoenolpyruvate carboxykinase
PEPCK-M	Mitochondrial form of Phosphoenolpyruvate carboxykinase
Per	Period
Pit1	Pituitary-specific transcription factor
pos	Position
QT	Quantitative trait loci
RFC	Residual feed consumption
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism
SPG	Specific gravity
SSCP	Single strand conformation polymorphism
SSLP	Simple sequence length polymorphisms
T	Thymine

CHAPTER 1

INTRODUCTION

New methods in molecular biology have made it possible to localize and characterize genetic loci associated with complex traits at the DNA level. This provides a new dimension for breeding of healthier and more productive animals. DNA plays an essential role. It is the master molecule that contains the information necessary to propagate cells, organs and the organism itself. In addition to selection, information at DNA level also enables breeders to develop superior hybrids by combining genotypes associated with overdominance (epistasis). Furthermore, it contributes to our understanding of the basic genetic architecture and reads the imprints evolution has left in the genome. Often traditional approaches have been unsuccessful in identifying genes influencing complex or quantitative traits. Such traits are determined by many different interacting genes and do not display simple Mendelian patterns of inheritance.

Two approaches have been mostly used to address the genotype-phenotype relationship. The method which has received initially most attention is the mapping of quantitative trait loci (QTL) by determining the co-inheritance of phenotypes and marker genotype in several generation. It is an extension of Mendel's experiments with peas. Once the chromosomal location of a QTL is known, the region can be analyzed for the genes that are potentially responsible for the phenotypic variation. This strategy has not been successful for complex traits that involve multiple genes and gene-environment interactions (Doris, 2002; Rannala, 2001; Devlin *et al.*, 2001). The fundamental problem

is that in complex traits the one-to-one relationship between phenotype and genotype breaks down (Lander and Schork, 1994).

An alternative approach is an epidemiological approach, searching for association between markers and traits in a randomly mated population. One common strategy is the candidate gene approach, which tests the association of traits with variants in known genes. Candidate genes can be screened relatively quickly and inexpensively. The approach is limited to known genes. However, the number of species whose entire genome is sequenced is increasing and their functional genes and variants are being catalogued. Sequencing of the chicken genome is underway and should be completed in the spring of 2004. Undoubtedly the bovine and porcine sequence will soon follow. In parallel to the establishment of reference sequences, comprehensive catalogues of DNA polymorphisms are being established. In the human genome project, many individuals provided DNA. This led to the identification of millions of DNA polymorphisms (Mullikin *et al.*, 2000; Altshuler *et al.*, 2000). The goal is now to establish a comprehensive catalogue of all polymorphisms where the frequency of the minor allele exceeds 1%. Towards this goal new high throughput technologies are being developed.

DNA polymorphisms can be grouped into haplotypes (i.e. arrays of DNA polymorphisms which co-segregate in the population of interest). This classification greatly reduces the genotyping required to characterize the genetic variation in a particular genomic region, often by using a few diagnostic DNA restriction fragment length polymorphisms (RFLP). It can also be used to construct phylogenic trees of particular chromosomal segments, trace the evolution of genes and deduce the presence of previous genetic bottlenecks (i.e. selection pressure during domestication of farm animals). A typical application of such knowledge is to conduct association analyses

between traits and genotypes determined by the genetically most distant haplotypes. Such genotypes are expected to have the most pronounced effect on phenotypes.

Chickens are the major source of animal protein in North America and consumption of poultry products is increasing rapidly in developing countries. Of major interest to the poultry industry is the selection for production traits and disease resistance. Phenotypic selection will inevitably affect many different genes. Selection at the DNA level provides a more targeted approach, since only a single gene is affected. However, it requires that the phenotypic consequences of such selection are well studied before application at an industrial scale.

In our laboratory, we have identified variants in several candidate genes and studied their association with production traits. Previous works using single markers in genes of the growth hormone axis and energy metabolism revealed significant associations for several genes. Preliminary results indicated that phosphoenolpyruvate carboxykinase (PEPCK) (Li *et al.*, 1998a) and ornithine decarboxylase (ODC) (Aggrey *et al.*, 1996) segregated for alleles that had strong effects on production traits.

PEPCK is a major regulatory enzyme in gluconeogenesis. It catalyzes the formation of phosphoenolpyruvate (PEP) by decarboxylation of oxaloacetate while hydrolyzing GTP. In vertebrates there are two isozyme forms, a cytosolic form (PEPCK-C) and a form present in the matrix of the mitochondria (PEPCK-M) (Hanson and Reshef, 1997). About 4 kb of the promoter region of the PEPCK-C has been sequenced (Sato *et al.*, 1997).

ODC synthesizes putrescine from ornithine, which along with spermidine and spermine (polyamines) are essential for DNA synthesis and cell proliferation (Fredlund and Oredsson, 1996). Around 5 kb of genomic DNA of ODC gene in a broiler chicken

has been sequenced (Zhang, 1994). Preliminary studies on two RFLP markers in White Leghorn chicken showed ODC association with immune response traits (Aggrey *et al.*, 1996).

In summary, the hypothesis of this study was, although major regulatory genes such as PEPCK-C and ODC were under heavy selection during domestication, they still segregate for variants and affect metabolic traits.

The objectives of this study were to conduct an in depth analysis of these two genes in White Leghorn chickens. It comprised the following main components:

(1) Identification of DNA variants in the PEPCK-C and the ODC genes.

(2) Identification of the haplotypes and establishing a phylogenetic tree.

(3) Identification of restriction fragment length polymorphism that are diagnostic for different branches of the phylogenetic tree and can be used for rapid genotyping.

(4) Conduct a comparative analysis of the haplotype structure of the two genes to gain evidence for differential selective pressure during the domestication of the chicken.

(5) Association analyses between haplotypes and economically important traits.

(6) Analyze whether single or multiple mutations affected the traits.

(7) Provide evidence for gene interaction.

CHAPTER 2

LITERATURE REVIEW

2.1. IDENTIFICATION OF QUANTITATIVE TRAIT LOCI

Quantitative traits, such as weight and height show a continuous distribution of phenotypic values rather than the discrete values observed for a qualitative trait. Such traits are usually controlled by multiple genes and influenced by environmental factors. Chromosomal loci which harbour genetic variants that affect quantitative traits are called quantitative trait loci (QTL). Genetic loci that affect qualitative traits leading to discrete distributions (such as disease status) are often also called QTL, and we will use the term QTL also for such loci.

In the past two decades, numerous QTL that affect relatively simple Mendelian traits (i.e. traits determined by a single or few loci and hence highly heritable) have been identified by using genetic linkage methods. Although these methods have been successful in identifying QTL affecting relatively simple traits, they have not been successful to genetically dissect more complex traits that are determined by many different genes. The failure is the result of several features of complex traits. Particularly, most complex traits are determined by the co-inheritance of variants at several genes. As a result with the effect of a variant in one gene is conditional on the type of allelic variations in other genes (Kuhnlein *et al.* 2003; Tabor *et al.*, 2002; Hartman *et al.*, 2001).

QTL mapping by linkage analysis is based on measuring the coinheritance of markers and phenotype over a few generations. In human genetics, this method has now been largely replaced by an epidemiological approach that entails association analysis between of genetic markers and phenotypes in large random populations.

2.1.1. Mapping of quantitative trait loci (QTL) by genetic linkage analysis

The discovery of linkage analysis dates back to 1909, when Weinberg (reviewed by Xiong and Guo, 1997) noted that, when alleles at some loci are followed over several generations of mating, they approach random association only asymptotically. This asymptotic behaviour is related to the frequency of recombination. Loci that are proximate to each other need many more rounds of recombination to become randomised than loci that are distant to each other. Morgan and colleagues in the first two decades of the twentieth century exploited such linkage to localise genes relative to each other on chromosomes. It led to the development of a gene map of the fruit fly *Drosophila melanogaster* (Morgan, 1919). In honour of Morgan, distances between markers are measured in centi-Morgans, where 1 cM indicates an average of one recombination events between the two markers in 100 meioses.

Complex mathematical methods have been devised to localise regions of the genome that segregate for alleles that affect a quantitative trait and to estimate their contribution to the total variation of the trait. It is based on the analysis of crosses between two individuals (or homozygous lines bred) that differ for the quantitative trait of interest and a series of marker loci. The F_1 generation resulting from this mating is heterozygous at the putative QTL as well as at the marker loci. The gametes produced by the F1 generation will have undergone one round of recombination. The consequences of the recombination events are displayed by the offspring produced in a second round of mating (i.e. the F2 generation). Marker genotypes and trait values are determined in the F2 generation. If there is a putative QTL closely linked to a marker gene, they will co-segregate; if they are distant from each other they will randomly segregate. Hence a

comparison of the phenotype and marker genotype can be used to locate QTL (Lander and Botstein, 1989)

An alternative mating strategy is to backcross F1 individuals with individuals of one of the homozygous parent populations. In this case only the gametes of the F1 parent have segregated for recombinants. It simplifies the interpretation of genetic data since the marker genotypes of offspring are diagnostic of the recombination products formed by only one of the parents. Backcross mating strategies have been most commonly used in QTL mapping studies. An example of a QTL scan in chickens is the search for Marek's Diesease resistance (MD) loci by Vallejo *et al.* (1998). They conducted a backcross starting from two homozygous lines of chickens that had been divergently selected for resistance to MD virus and identified five QTL associated with susceptibility.

Since chromosomal location of the genetic markers will indicate the chromosomal location of the QTL, much effort has been spent to establish linkage maps of markers in many species, including the major farm animals. Publicly accessible maps for many farm animals are available at the web-site "http://www.thearkdb.org" and include the pig, bovine and chicken. The site lists the map position as well as the primers and conditions to type these markers by the polymerase chain reaction (PCR). Linkage maps covering the entire genome make it possible to conduct a comprehensive analysis of all QTL and to estimate their relative contribution to the trait differences between the original parental strains.

There are a number of limitations to linkage analysis. In humans, where linkage mapping is based on pedigrees rather than planned matings, the number of individuals is often too small. For some animal species where matings can be carried out, the approach is time-consuming for animals with long generation times. A more serious problem stems

from the genetic architecture of genetic traits. The statistical models used in linkage analysis assume a cumulative additive effect of discrete loci of several major loci and by polygenes (i.e. the sum of small effects caused by many different genes) (Abecasis *et al.* 2000). However, there are many examples from humans, mice and chickens which show that the majority of traits are determined by the interaction of several genes. Due to such interactions, the effect of the genotype at one locus may be entirely dependent on the genotypic status of another gene (Kajiwara *et al.*, 1994; Feng *et al.*, 1997; Nadeau, 2001). In such cases, methods based on linear models may fail to detect many QTL.

A dependence of the genotypic effect on the genetic background has been observed in knock-out mice, where the consequences of inactivating a gene is often strain dependent (Nadeau, 2001). In a non-inbred strain, each offspring has its own genetic background due to the random assortment of the parental genes. Hence, in order to gain insight into the overall effect of a QTL many different inbred homozygous lines have to be generated and crosses conducted and the spectrum of QTL may differ from line to line.

Even when linkage has been identified, it has seldom led to the resolution of linkage regions to less than a few centimorgans (cM) in most complex traits (Weiss and Terwilliger, 2000). One cM corresponds to a physical distance of about 1 million base pairs. Although genes in such a region can be identified, especially when the entire genome sequence is known (sequencing of chicken genome is now underway), it is still an immense task to identify the gene and the particular alleles responsible for the trait variation. At the end it requires a database amenable to the type of association analysis described in the next section.

It has been proposed that marker assisted selection (i.e. selection for the markers used to map the QTL) may be used as an adjunct to selection at the phenotypic level.

However, the phase between marker and putative QTL (i.e. gametic linkage between marker and QTL allele) may differ in different parents (or homozygous inbred lines). The gain achieved by selecting for a favourable marker allele identified by linkage analysis is therefore limited.

2.1.2. Identification of QTL by association analysis

An alternative to QTL mapping by linkage analysis is association analysis (sometimes called linkage disequilibrium mapping or LD mapping). Rather than following the co-inheritance of markers and phenotypes in a few generations, this analysis searches for associations between markers and phenotypes in a population that had been randomly mated over many generations. The rational is that in such a population only markers and QTL in close proximity are associated (at linkage disequilibrium) whereas linkage between markers and QTL that are more distant from each other has decayed due to recombination.

Although initially frowned upon, Risch and Merikangas (1996) indicated that association analysis may statistically be more powerful than linkage analysis. However, it is critically dependant on the genetic structure of the population group being analysed. In particular when searching for QTL associated with human diseases the control group and the affected group have to be carefully matched in order to avoid admixture of population with different genetic backgrounds (Wacholder *et al.* 2000; Wacholder *et al.* 2002).

Chicken strains are ideal for association analysis, since strains are available that have been kept as closed breeding populations for many generations. Association between markers can then be established by comparing the marker frequency among individuals with trait values below the median with the marker frequency among individuals above

the median. More extreme, one may compare the frequencies of markers among individuals from opposite tails of the trait distribution (Kuhnlein *et al.*, 1997; Feng *et al.*, 1998). Care has to be taken that the effective population size is large; otherwise individuals within groups may be more closely related than individuals between groups due to chance. This may lead to false association due to the inheritance of allele genes unrelated to the trait of interest.

An alternative method is to regress the trait values on genotypes. This type of analysis is most commonly used and has received extensive mathematical treatments (Abecasis *et al.*, 2000; Kennedy *et al.*, 1992). In particular, relatedness between individuals is accounted for by including a correction term for a polygenic (the cumulative effect of many genes with small effects). This type of analysis assumes additivity between QTL and the polygenic background, a model which based on our current knowledge of biology is difficult to defend. Simulations indicate that the effective population size and sample sizes have to be quite small in order to see such effects (Kennedy *et al.*, 1992).

Chickens have a relatively short generation time and are kept as closed populations. Such strains can be divergently selected for the trait of interest and the influence of such selection on allelic frequencies can be determined. Since marker frequencies may be affected by genetic drift, selection has to be carried out in strains with a large effective population size or in duplicate strains. This approach has been used to provide evidence for co-selection of alleles of the GH gene and phosphoenolpyruvate carboxykinase-M (PEPCK-M) with MD resistance (Kuhnlein *et al.*, 1997; Li *et al.*, 1998a).

In human genetics, association analysis has become the method of choice. The number of evenly spaced genetic markers to cover the entire human genome has been estimated to be 500, 000 (Kruglyak, 1999). This poses a problem in terms of labour, DNA requirement and the power of the statistical analysis due to the multiplicity of tests. In order to reduce the number of tests, Collins et al. (1997) proposed two general strategies, one direct and one indirect. Both rely on the hypothesis that common genetic variants underlie susceptibility to common diseases (or complex traits). The direct strategy is to restrict association analysis to variants in coding and regulatory regions of genes. It is assumed that such polymorphism maybe directly affect gene function and be responsible for trait differences. The problem is that we are far from being able to recognize DNA sequences that are functional (see for example the review by Eddy (2001) on non-coding RNA genes).

The indirect strategy avoids the problem of identifying polymorphisms in regulatory and coding sequences. Instead, alleles in genes are grouped into haplotype (alleles which are co-inherited). As described later, a few genetic markers are often diagnostic for different haplotypes, thus greatly reducing typing and the multiplicity of association tests (Tabor et al., 2002; Lewontin, 1988). Significant associations can be used to predict molecular or biochemical mechanisms for the sequence variants. Experiments can then be designed to test these putative functional roles in biological processes. Such follow-up experiments are necessary, since some associations may arise when the linkage disequilibrium (LD) between the marker and QTL extends to several genes located nearby (Kruglyak, 1999).

2.1.3 The candidate gene approach

Although it is likely that in the future high throughput methods to measure diagnostic DNA variants will be developed, current analyses are carried out with a few genes at a time. Genes in such an analysis are generally selected on the basis of their biological properties, hence they are called functional candidate genes. Examples in human genetics are the collection of sequence variants in 75 candidate genes for high blood pressure by Halushaka *et al.* (1999) and of 106 candidate genes for heart disease by Cargill *et al.* (1999). Alternatively, one may select genes on the basis of prior QTL mapping (positional candidate gene) or association analyses in other data sets.

The candidate gene approach was first applied in 1952, when Linus Pauling hypothesized that an altered sickle hemoglobin may be the cause rather than the consequence of sickle cell anemia (Johnson, 1984). This was confirmed by analyzing the amino acid sequence of the hemoglobin in affected and unaffected individuals. This revealed a change in a hydrophobic amino acid to a hydrophilic amino acid which altered the physical properties of hemoglobin structure and hence the red blood cell function (Linde, 1972).

Since the identification of this disease causing mutation, the identifying of mutations affecting disease susceptibility have expanded and at least once a week the discovery of a new "disease gene" is reported. At the early going, most genetic disorders characterized at the DNA level were so called Mendelian traits, where a single gene was responsible. This is now rapidly shifting to the identification of genes contributing to more complex diseases which result from the combination of alterations in many different genes.

The candidate gene strategy has also been extensively used to identify quantitative trait gene in farm animals. Traits of interest are related to production and hence most often related to growth, energy metabolism and disease resistance. Disease resistance is also an issue in relation to food safety. In cattle, the influence of genetic variants in kappa casein gene (Ng-Kwai-Hang *et al.*, 1991), growth hormone (Yao *et al.*, 1996, Falaki *et al.*, 1996), growth hormone receptor (Falaki *et al.*, 1996), ornithine decarboxylase (Yao *et al.*, 1995, 1998) on milk production and composition were studied. In chickens, the candidate genes mostly studied in our laboratory were genes of the GH axis including GH, GHR and IGF-I (Kuhnlein *et al.*, 1997; Feng *et al.*, 1997,1998; Nagaraja *et al.*, 2000) and mitochondria (Li *et al.*, 1998b) and PEPCK-M (Li *et al.*, 1998a).

The candidate gene approach investigates whether a known gene segregates for variants that affect a trait. Hence it precludes the discovery of new genes. To partially circumvent this problem, one may conduct mRNA expression analyses and identify those genes whose expression correlates with the trait of interest. A specific example is the identification of host genes which were induced in chicken embryo fibroblasts following infection by Marek's disease virus by Morgan *et al.* (2001).

Similarly, Li *et al.* (1998c) devised a pooling method to identify genes segregating for markers whose frequency was affected by divergent selection. It was conducted by a bulk analysis of pooled genomic DNA from different strains for RFLP using anonymous liver cDNA clones as probes. The cDNA clones which revealed genes that were co-selected were then sequenced and identified by searching of homology with sequences in chickens and other organisms (Li *et al.*, 1998c). This same approach was used to identify genes by RFLP that differ in their frequency among the individuals at the top and tail of a trait distribution.

2.2. DNA polymorphism AND haplotypes

Since the introduction of molecular methods for the study of DNA sequence variation, very large numbers of variable nucleotide positions have been discovered in a great variety of organisms. This molecular variation includes both single nucleotide polymorphisms (SNPs) as well as insertions and deletions of longer nucleotide sequences.

The human genome project, as an additional benefit to the sequence itself, yielded data on the types and frequencies of mutations covering the entire human genome. Ventner *et al.* (2001) catalogued 2.1 million polymorphisms, while the International SNP Map Working Group (2001) catalogued 1.4 million DNA polymorphisms. The latter data set is now accessible on the internet. An effort is now underway to identify all mutations whose minor allele occurs at a frequency exceeding 1% in the entire human population. Such compilations will undoubtedly also become available for other species, as their genomic DNA will be sequenced.

Among the mutations, single nucleotide exchanges are most frequent. Such exchanges may be the replacement of one pyrimidine against another (transition) or a pyrimidine for a purine (transversions). Transitions are about 10 times more frequent than transversions. The probability that a given nucleotide undergoes a change depends on many factors, including the chemical stability of nucleotides, the chromatin structure, the fidelity of DNA replication and the accessibility to DNA repair. A particularly frequent event is the transition of a C by a T when the C is followed by a G. This reflects that cytosine methylation only occurs when C is followed by a G. When methylated cytosine is deaminated it yields thymidine, which may escape mismatch repair and become fixed in the genome (Hacia *et al.*, 1999).

Sites of deletion and insertions are less frequent. Among these, simple sequence length polymorphisms (SSLP; i.e., microsatellites, variable tandem repeat loci etc.) play a particularly important role. As its name implies, alleles at these loci arise from a variable number of repeats, due to slippage during DNA replication or gene conversion betweeen SSLP loci of the same type. SSLP loci are less frequent then SNPs loci, but their mutation rate is one or two orders of magnitude higher. As a consequence, two individuals differ at many SSLP loci, even if they are closely related. This property has made them useful for determining relatedness among individuals in forensic medicine and population studies of wildlife.

In the early going, SSLP loci have played an important role as genetic markers to establish gene maps and to determine the location of QTL (Witmer et al., 2003). With the current shift to the comprehensive analysis of all mutations of a genome, the role of SSLPs in identifying quantitative trait mutations is diminishing.

2.2.1 Detection of DNA polymorphisms

A comprehensive catalogue of DNA polymorphisms leaves the researcher with the problem of identifying the alleles of known polymorphisms in all individuals of a particular population of interest.

One of the first methods used was the PCR-RFLP method. It involves the PCR amplification of the DNA segment of interest, followed by digestion of the PCR products by diagnostic restriction enzymes and sizing of the cleavage products by gel electrophoresis. This method is restricted to DNA variants which alter the restriction enzyme cleavage pattern of the DNA segment (usually an SNP in the recognition site of the restriction enzyme). Alternatively, a change of the DNA sequence may alter the

single-strand conformation of the amplified DNA segment, which may be recognizable by electrophoresis through a denaturing gel (SSCP). These two methods are still the most frequently used techniques for the analysis of a small number of samples.

New methodologies are developed which can handle the analysis of hundreds of thousands of SNPs in tens of thousands of individuals (Chicurel, 2001). Three of the many methods proposed are the hybridization assay, the primer extension assay and the invader assay. In the hybridization detection system, the sample DNA is tagged by a chromatophor and hybridized to SNP specific oligonucleotides immobilized on chips. SNPs can then be recognized scanning the DNA chip for differences in the hybridization pattern. The difficulty with this methodology is to find conditions or design primers that simultaneously detect hundreds or thousand of different SNPs in a hybridization reaction (Fan *et al.*, 2000).

In the primer extension assay, primers are designed to hybridize adjacent to an SNP. Only primers without mismatch at their terminal 3' nucleotides are extended by DNA polymerase and extension products can be measured by providing chemically tagged nucleotides that can be identified by fluorescence or by mass-spectrometry.

A more recent method is the 'invader assay' (Lyamichev *et al.*, 1999). This assay makes use of a DNA repair enzyme that recognizes the 5' flap that can be created by hybridyzing two overlapping primers to a target DNA. Cleavage of the flap is dependent on a perfect match 5' to the cleavage site, thus providing an assay for SNPs. Ohnishi et al. (2001) have designed a protocol that combines the 'invader assay' with multiplex PCR to provide sufficient quantities of sample DNA. They claim that with their current methodology they can achieve a throughput of 100 million SNPs per year.

As indicated above, these assays are designed to measure known SNPs. Other methods target the methodologies of DNA sequencing and aim at re-sequencing an entire genome in a few days (Adam, 2001).

2.2.2 Genomic distribution of DNA polymorphisms

The induction of mutations is not random, but depends on the interplay between proof reading, DNA repair mechanisms, chromatin structure and the DNA sequence itself. An even larger factor is the biological constraint. Mutations that impair the reproductive success of an individual will rapidly disappear from the species. Critical regions for survival are coding sequences and sequences that regulate transcription or the processing of RNA transcripts. Hence, within a species mutations in such regions are rare and the mutation profile can be used to predict whether a region is in a coding or a regulatory region.

Dermitzakis et al. (2002) have extended this type of analysis to a comparison of different species. They conducted a sequence comparison of human chromosome 21 (33.5 megabases) with the syntenic regions in the mouse. Sequence comparison revealed about 3500 blocks > 100 bp with a sequence identity> 70%. As expected, many of the conserved blocks contained exonic sequences. However, about two thirds of the blocks were non-genic. It will be exciting to unravel their potential biological function.

The spectrum of polymorphisms is not only affected by short-term functional constraints, but also by variants that affect fitness in the long-term. The frequency of a variant that increases survival of a species will increase in the population. This will lead to the co-selection of neutral marker alleles that are in phase of the variant allele providing a selective advantage. As a consequence, markers within or adjacent to genes

under selective pressure will be at a linkage disequilibrium (non-random association of alleles).

2.2.3 Haplotype blocks and linkage disequilibrium

Chromosomal regions of linkage disequilibrium are called haplotype blocks and the arrays of marker alleles that are observed are called haplotypes (Wall and Pritchard, 2003). The size of such blocks is determined by the evolutionary history of a population and the frequency of recombination that leads to the randomization of markers. The identification of such haplotypes is useful, since they can often be characterized by a small number of DNA polymorphisms, thus reducing the number of DNA variants that have to be typed in association analyses of traits. The United States National Human Genome Research Institute has therefore initiated the HapMap Project. Its goal is to create a genome map of markers in linkage disequilibrium.

Analysis of haplotypes in humans revealed distinct regions of high linkage-disequilibrium (LD) or haplotype blocks, interspersed with regions of low LD between markers. Analysis of a 500 kb region of the human chromosome region 5q31 revealed haplotype blocks of 3-92 kb (Daly *et al.*, 2001). In these blocks only 2 to 4 haplotypes were observed. Even longer blocks were observed in other studies (Dawson et al., 2002).

The forces with create the haplotype-block patterns are complex. Some but not all of the boundaries of these blocks are due to an uneven distribution of recombination events. Hot-spots of recombination are often found at the boundaries of blocks and other areas of the chromosome may exist where recombination is suppressed. Other events that contribute to the creation of haplotypes may be genetic bottlenecks and/or recent selection pressure. As an example, Tishkoff *et al.* (2001) conducted a haplotype analysis of the

glucose-6-phosphate dehyrogenase (G6PD), a gene which influences susceptibility to malaria. The analysis indicated that malaria had a major impact on human populations and led to the predominance of two haplotypes in populations originating from areas with a high incidence of malaria. Both of these haplotypes are associated with low G6PD activity.

Haplotype information can be used to construct a gene tree that provides further information of the evolutionary history of a gene. In the absence of recombination or gene conversion, building such a tree within a species is straight forward. Consider two mutations with alleles A, a and B, b, respectively. Assuming that the mutation from B to b has only occurred once in evolution and is the most recent, it must have occurred on allele A or a, but not on both. If it is on allele A, only the combination (A-B), (A-b) and (a-B) will be observed, but not the combination (a-b) (Note: since one of the four combinations is missing, the markers are at maximal linkage disequilibrium). Such considerations will lead to an unambiguous unrooted tree. In the examples we present in this thesis there are no contradictions and the tree coincides with every tree obtained by using any of a series of computer programs. If the tree is rooted using an extant species, the tree also yields the flow direction of the mutation event at each site.

Phylogenetic trees identify the evolutionary distance between haplotypes. This information is useful when a limited sample size requires pooling of genotypic classes for QTL analysis by association. Based on the working hypothesis that a quantitative trait mutation only occurred once in evolution, one may compare genotypes defined by the haplotypes segregating at a given branch point. By moving through the branch points one may pinpoint branches in which the quantitative trait mutation has occurred, even when

the QTL mutation itself has not been identified. This analysis also addresses the question whether a gene harbors a single or multiple QTL.

2.3. EPISTASIS AND BUFFERING

Complex traits are determined by networks of interacting genes (Kuhnlein *et al.*, 2003; Hartman *et al.*, 2001). The potential number of gene interactions (two-locus, three locus, etc) is enormous. Such networks are overlapping and may have evolved to buffer the phenotype of an organism against variations caused by the environment. The importance of such buffering has been fully recognized and it is thought that many human diseases are the consequences of a breakdown of homeostatic process due to the co-inheritance of mutations in several different genes (Wanstrat and Wakeland, 2001). In such a model, effects of quantitative trait mutations are interdependent and in the extreme case even conditional of each other.

A QTL is usually defined as a locus that affects the central tendency (median or mean) of a trait distribution. However, systems where traits are determined by a network of interacting genes, QTL are also expected to affect the variance of a trait distribution. In particular, in an outbred population which segregates for alleles at many other genes, a breakdown of the homoeostatis is expected to increase the standard deviation of a trait and may have little or no effect on the mean or median. Although tests for unequal variances are available, their statistical power is largely unknown.

An additional problem stems from the non-additivity of the genotype and the physiological status of an individual. Specifically, the physiology of an indvidual at one end of a trait distribution will differ from the physiology of an individual at the other end of the trait distribution. Hence, gene replacement may affect individuals from the two

tails of a distribution differently. As a consequence a gene may affect the shape of the trait distribution (usually a change in the median as well as the variance)

Homeostatic mechanisms are also evident at the level of trait correlations. An example is the analysis by Nagaraja *et al.* (2000) who analyzed the partial correlation coefficients between the rate of egg-laying, egg weight, body weight and feed consumption in a strain of White Leghorn chickens. All four traits were significantly correlated, ensuring that a chicken lays eggs at a size and frequency commensurable with its body weight and feed intake. These researchers found that markers in genes of the GH-axis affected the correlation between traits, emphasizing again the difficulties to adequately define QTL.

Theoretically, all possible two-way, three-way, and up to *n*-way interactions among QTL can be measured through *n*-way ANOVA or related statistics. Practically, there are several problems with this approach (Zeng *et al.* 1999; Merila and Sheldon, 1999). Firstly, segregating population sizes are usually too small for accurate estimates of multi-locus interactions. To measure the interaction between two loci, all possible genotypes for those loci must appear in the population in sufficient frequencies to allow statistical comparisons. Taking the simple case of two independent QTL segregating in a closed population, there would be nine possible two-locus genotypes. Three-way interactions would result in 27 genotype classes. To have a good measure of the phenotypic effect of each class (and thus to gain a more accurate measure of epistasis) would require a very large population that most of the time is not available.

Second, the number of potential multi-locus interactions is very large, requiring many statistical tests, some of which will, by chance, reach statistical significance. One

solution is to raise the significance threshold to avoid reporting spurious interactions, hence requiring an additional increase in sample size.

Because of the limitations listed above, it seems unlikely that it will be possible to produce detailed descriptions of QTL by QTL interactions using primary segregating populations. More likely, it will be necessary to construct nearly isogenic lines with single QTL and combinations of QTL that can be replicated in experimental designs (such as segregation analysis within families) to allow a more precise measurement and description of interactions.

2.4. AGRICULTURAL ANIMALS: A DIFFERENT PARADIGM?

Most of the arguments presented above were developed with the interest of identifying the genetic contribution to disease in humans with the aim of understanding the underlying mechanisms at the molecular level and to devise methods to intervene in the disease process or recommend preventative measures. The goal in farm animals is different. Aberrant phenotypes, rather than being the subject of intensive analysis, are eliminated from the breeding stock. The interest then is to further improve strains by selection at the DNA level rather than understanding the molecular underpinning of a trait (Andersson *et al.*, 2001).

A constraint is that most breeders are hesitant to select for rare genes. In poultry, pure strains are kept as closed breeding populations under constant selection. The pure lines are then intercrossed to yield the commercial product. Introgression of a gene into the pure lines may disrupt the genetic fabric of the primary breeding stock which had been the result of careful selection over many years. The interest then is generally to

identify genetic variants of interest in the existing strains and to use marker assisted selection as an adjunct to selection at the DNA level.

With this goal in mind, the focus should be in conducting association analysis between traits and genetic variants that are relatively frequent in the strain of interest. This may be more fruitful than breeding strains of extreme phenotypes and conducting segregation analysis between markers and traits. Further, since hybrids are the final product, the emphasis should be on gene interaction.

The goal of QTL mapping and association analysis is to identify a QTL that affects a particular trait. One may reverse the question and ask whether a particular gene segregates for alleles that have biological effects. As an example, it is known that the GH-axis affects many traits such as growth, milk production in the bovine, as well as the immune system. Hence, if a strain segregates for variants in genes of the GH-axis that affect spatial and temporal expression, they should affect major production traits. Once the existence of such variants has been recognized, the analysis can be extended to other traits and trait correlations to provide predictors for the outcome of future marker assisted selection.

2.5. GENES ANALYZED IN THIS THESIS

In our laboratory, my previous colleagues have analyzed genes in the GH-axis, a pathways known for its importance in many aspects of development and metabolism. In this thesis, I focused on two other genes which are pivotal in development and homeostasis. The genes were phosphoenolpyruvate carboxykinase (PEPCK), a major gene in the regulation of gluconeogenesis and ornithine decarboxylase (ODC), an important regulatory gene of cell division, differentiation and apoptosis.

2.5.1 Phosphoenolpyruvate carboxykinase (PEPCK)

PEPCK plays a critical role in gluconeogenesis, the net synthesis or formation of glucose from a large variety of non-carbohydrate substrate. The enzyme uses a variety of substrates including amino acids, lactate, pyruvate, propionate and glycerol as sources of carbon. Two major cycles between tissues have identified in gluconeogenesis, the Cori cycle and the alanine cycle.

Lactate is the source of carbon atoms for gluconeogenesis via the Cori cycle. During anaerobic glycolysis in skeletal muscle, pyruvate is reduced to lactate by lactate dehydrogenase (LDH). The lactate is released to the blood stream and transported to the liver where it is converted to glucose. The glucose is then returned to the blood for use by muscle as an energy source and to replenish glycogen stores.

The alanine cycle is used primarily as a mechanism for skeletal muscle to eliminate nitrogen while replenishing its energy supply. Pyruvate produced by glycolysis can accept amino groups from amino acids to form alanine. This reaction is catalyzed by glutamate-pyruvate transaminase, GPT (also called alanine transaminase). Additionally, during periods of fasting, skeletal muscle protein is degraded for the energy value of the amino acid carbons and alanine is a major amino acid in protein. The alanine then enters the blood stream and is transported to the liver. Within the liver alanine is converted back to pyruvate which is then a source of carbon atoms for gluconeogenesis. The amino group transported from the muscle to the liver in the form of alanine is converted to urea in the urea cycle and excreted.

One of the intermediates in gluconeogenesis is glycerol-1-phosphate. It had been proposed 30 years ago that glycerol-1-phosphate drives the re-esterification of free fatty acids (Ballard *et al.*, 1967). In particular, fat cells lack the last two enzymes of

gluconeogenesis. In these cells, glycerol-1-phosphate is the end product. It reacts with fatty acids released from lipids and prevents their release into the blood stream. Genetically engineered mice that do not express PEPCK in fat cells are lean, whereas mice that over express PEPCK show an increase in fat deposition (Olswang *et al.*, 2002)

Gluconeogenesis reverses the sequence of reactions of glycolysis. Most of the steps leading to the synthesis of glucose from pyruvate are reversible and controlled by the same enzymes as glycolysis. However, three glycolytic reactions are irreversible and are bypassed by reactions requiring separate enzymes. These irreversible reactions (or series of reactions) are the formation of phosphoenolpyruvate from pyruvate, the formation of fructose 6-phosphate from fructose 1,6-disphosphate and the dephosphorylation of glucose 6-phosphate. These bypass reactions are the major regulatory sites of glucose metabolism. PEPCK, the enzyme analyzed here catalyzes the formation of phosphoenolpyruvate (PEP) by decarboxylation of oxaloacetate while hydrolyzing GTP.

$$\text{Oxaloacetate} + \text{GTP} \longrightarrow \text{PEP} + \text{GDP} + CO_2$$

The enzyme was first identified and characterized by Utter and Kurahashi (1953). The protein occurs in two isozyme forms in vertebrates, a cytosolic form (PEPCK-C) and a form present in the matrix of the mitochondria (PEPCK-M). The two forms of PEPCK are distinct proteins, which are distinguishable by their immunochemical properties and by their isoelectric points. The two isoenzymes are encoded by separate mRNA species, indicating that they are synthesized independently of each other and encoded by separate nuclear genes, yet have similar kinetic properties and approximately the same molecular weight (Hanson & Patel, 1994). In mammals, the cytosolic form of PEPCK (PEPCK-C)

accounts for over 95% of the PEPCK activity in the liver and kidney of the rat and research has therefore almost exclusively focused on this enzyme.

The level of PEPCK is controlled by a wide variety of physiological stimuli including hormones, dietary carbohydrate, and cellular intermediates (Hanson *et al* 1994, Hanson & Reshef, 1997). Conversely, analysis of transgenic mice indicates that modulating the expression of PEPCK-C in the liver leads to changes in the level of many enzymes involved in energy metabolism (She et al., 2000). Hence, the major role of PEPCK may be the regulation of hepatic energy metabolism, rather than gluconeogenesis. PEPCK-C has a short half-life of only 6-8 h (Hanson & Reshef, 1997), but factors that regulate its rate of degradation have not yet been discovered.

In chickens, carbohydrate metabolism differs from mammals. This is partially due to differences in PEPCK expression and the relative contribution of the cytosolic and mitochondrial isoforms. In chickens, the kidney contributes up to 30% of gluconeogenesis and is the major regulatory component, whereas, in mammals the contribution of the kidney to total gluconeogenesis is negligible and the major controlling component is the liver. Further, in chickens PEPCK-M and not PEPCK-C is the major isozyme, accounting for 50% to 80% of the activity in the kidney (Bisbis et al., 1994). Further, Savon et al. (1993) reported that hepatic PEPCK-C in mammals is only expressed at birth, while in chickens it is expressed throughout embryonic development and is negligible at the time of hatching.

The c-DNA of PEPCK-C has been cloned from many organisms (Hod *et al.*, 1984; Cook *et al.*, 1986). Rat PEPCK-C with molecular weight of 69,289 contains 621 amino acids (Beale *et al.,* 1985). The same isozyme from the chicken consists of 622

amino acids with the molecular weight of 69,522, and shares 80% amino acid identity with rat PEPCK-C (Cook *et al.*, 1986).

The promoter of the rat PEPCK-C gene has been studied intensively. At least a dozen regulatory elements have been identified in the 5'-flanking region of the PEPCK-C gene (Hanson & Reshef, 1997). The nucleotide sequence of the promoter and its flanking regions (nucleotide −1855 to +2083) in the chicken PEPCK-C gene was determined by Sato et al (1997). The nucleotide sequence of exons 1, 2, and 3 were highly homologous to the corresponding exons of the rat gene. In contrast to the mammalian PEPCK-C genes that have been studied, the promoter of the PEPCK-C gene from the chicken varies in its arrangement of regulatory elements, reflecting the very different role that PEPCK-C plays in avian glucose homeostasis (Savon et al.,1997).

Regulation of glucose homeostasis in birds differs markedly from that in mammals. Homology of the sequence –1 to – 500 to that of rat gene was 52% and most of the hormone-responsive sequences in the rat gene, such as the glucocorticoid-responsive region, were not conserved in the chicken gene (Sato et al., 1997).

In comparison to PEPCK-C, the mitochondrial form of the enzyme from the chicken has a longer half-life (48 hours), and its synthesis is not under acute control (Hanson & Reshef, 1997). Chicken PEPCK-M has been cloned and it consists of 607 amino acids with molecular weight of 67,186. The central region of the protein (residues 200 – 400) shares about 80% amino acid identity with chicken PEPCK-C. Similarity decreases sharply at the amino and carboxy–terminal ends of the two proteins (Weldon *et al.*, 1990).

PEPCK-M is unusually polymorphic at Msp1 and Taq1 restriction sites (Li *et al.*, 1998a). Seven alleles were found in six pairs of chicken strains derived from different

genetic base populations, each pair differing in Marek's disease susceptibility. A comparison of strains divergently selected for resistance or susceptibility to Marek's disease revealed that some bands were responsive to selection. The frequency of one of the alleles was reported higher in the MD susceptible strains than the resistant strains (Li *et al.*, 1998b). The effect of variants of PEPCK- M on parameters of egg production in chicken has been analysed and age at first egg (AFE) was significantly correlated with an Acc*I* RFLP located in exon 9 (unpublished observation).

2.5.2. Ornithine decarboxylase (ODC)

The second gene analyzed, was ornithine decarboxylase (ODC). ODC catalyzes the decarboxylation of ornithine to form putrescine (Morgan, 1999). Putrescine is the precursor of spermidine and spermine. These polyamines are present in all living cells including prokaryotes, eukaryotes, plants and animals (Pegg & McCann, 1982). Polyamines are organic cations since the primary and secondary amino groups are protonated at a physiological pH. Putrescine is divalent, spermidine is trivalent and spermine is tetravalent. Polyamines are important modulators of DNA-protein and protein –protein interactions and hence are essential for life. They are the mediator of regulatory signals that affect cell division, apoptosis and cell differentiation and indeed, most signal pathways interact with polyamine biosynthetic pathways and the regulation of intracellular polyamine levels (Alm et al., 2000, Thomas and Thomas, 2001).

At the level of the organism, polyamines play important roles in embryonic development (Kusunoki & Yasumasu, 1978), cancer (Thomas & Thomas, 2003), neurochemistry (Seiler, 2000), as well as pulmonary (Hoet & Nemery, 2000) and immune system functions (Seiler & Atanassov, 1994; Thomas *et al.*, 1992). Cancer cells have

elevated levels of polyamine biosynthetic enzymes. In particular, high expression of ODC is mandatory for a malignant phenotype (Konno, 2001; da Rocha *et al.*, 2002; Thomas & Thomas, 2003). ODC has therefore become the target for the development of anticancer drugs. Agents that inactivate ODC irreversibly have proven to be powerful therapeutic agents to cure African sleeping sickness. The principle of action is that the ODC of the trypanosomes is particularly sensitive to these agents because of its slow turn-over. In human cells, inactivated ODC is rapidly replaced due to the rapid turn-over rate of the enzyme.

Organisms have numerous pathways to regulate polyamine levels. This includes the synthesis from amino acid precursors, the cellular uptake mechanisms that salvage polyamines from the diet and intestinal micro-organisms, as well as stepwise degradation and efflux (Alhonen *et al.*, 2000).

The genomic sequence of the human ODC has 11 introns (Fitzgerald and Flanagan, 1989). Further, in addition to the functional ODC gene, a pseudogene is also present (Hickok *et al.*, 1990). Johnson and Bulfield (1992) reported the cDNA sequence in chickens first. It was subsequently found that higher ODC mRNA levels and enzyme activity were associated with selection for rapid growth in a broiler line (Johnson *et al.*, 1995). This study confirmed the observation by Smith (1990) that dietary putrescine has a positive effect of the growth rate. In the latter study, White Leghorn chickens at one week of age were fed with different amounts of purified putrescine incorporated into the feed. The group feed 0.2% putrescine had significant increase of ODC activity and the fastest growth rate.

Zhang (1994) sequenced around 5 kb of genomic DNA of ODC in a broiler chicken. Two RFLP markers in White Leghorn in intronic sequences were identified and

preliminary studies indicated that they were associated with immune response traits (Aggrey *et al.*, 1996).

CHAPTER 3

GENETIC VARIABILITY OF THE CYTOSOLIC PHOSPHOENOLPYRUVATE CARBOXYKINASE GENE IN WHITE LEGHORN CHICKENS

R. Parsanejad, D. Zadworny and U. Kuhnlein[1]

Department of Animal Science, McGill University,

Ste. Anne de Bellevue, QC, Canada H9X 3V9

[1] To whom correspondence should be addressed: kuhnleinu@macdonald.mcgill.ca

Published by Poultry Science (2002. 81(11):1668-70)

3.1 ABSTRACT

Phosphoenolpyruvate carboxykinase (PEPCK) is a key regulatory enzyme of gluconeogenesis. Genetic variations in this gene may therefore affect a wide variety of traits, including tumor growth which is heavily dependent on glucose metabolism. We have previously shown the gene coding for the mitochondrial form of PEPCK (PEPCK-M) segregates for markers which are co-selected with resistance to Marek's disease. In this communication we analyzed the genetic variability of PEPCK-C, the gene which codes for the cytosolic form of PEPCK. A 3792 bp segment of 5'-region of the PEPCK-C gene (pos. -1723 to 2069) was sequenced in 4 individuals from 8 different strains of White Leghorn chickens (a total of 64 genomes). A total of 19 single nucleotide polymorphisms (SNPs) were identified. Neither deletions nor insertions were present. The most frequent SNPs were transitions (79%) and in most cases the ancestral allele coincided with CpG dinucleotides (approximatly10-fold by correcting dinucleotide frequencies). A gene tree was constructed assuming maximal parsimony. It led to the delineation of 6 haplotypes (combination of alleles). Two of the SNPs coincided with RFLP detectable by the restriction enzymes *Aci*I and *Bst*EII, respectively. Based on this analysis we can now identify individuals with the evolutionary most distant PEPCK-C haplotypes, establish strains of these haplotypes and analyze trait associations and epistasis with other genes.

3.2 INTRODUCTION

A survey of allelic variations in genes has several purposes. Firstly, haplotypes (i.e. combinations of allelic variations) can be delineated and their evolutionary relationship determined. Trait associations of genes can then be studied by segregation analysis between the most distant haplotypes. Further, strains can be selected for particular haplotypes to assess the consequences of marker assisted selection and to study the interaction with other genes. Secondly, programs aimed at conservation of genetic diversity of farm animals can be focused on strains which contain genetically distant haplotypes. In this communication we analyzed 3,792 base pairs of the 5' region of the gene encoding for the cytosolic form of phosphoenolpyruvate carboxykinase (GTP) [EC 4.1.1.32] (PEPCK-C), which is a key regulatory enzyme of gluconeogenesis (Hanson & Reshef, 1997).

3.3 MATERIALS AND METHODS

We analyzed DNA from 32 chickens of 8 different White Leghorn strains. One was an experimental strain (strain 7) established in 1958 from four commercial strains and kept without selection (Gowe *et al.*, 1993), whereas the other seven strains were different commercial strains. All strains had been kept as closed breeding populations for at least 20 generations. DNA was isolated from the blood of 4 chickens from each strain using standard procedures (Sambrook *et al.*, 1989). Four primer pairs were designed (DNAMAN[2] version 2.7) to cover the region from position -1,723 to 2,069 in four overlapping sections (Figure 3.1). The amplification protocol for each section was 5 min presoaking at 95 C, followed by 35 cycles of 60 s at 94 C, 80 s at 62 C and 90 s at 72 C. The PCR product was purified with a PCR purification kit[3] and was sequenced from each end using an ABI 310 capillary sequencer. In good quality runs, there was no difficulty in distinguishing heterozygotes from homozygotes.

3.4 RESULTS AND DISCUSSION

The 3,792 bp scanned contained 1,723 bp of the promoter region, 1,578 bp of introns, and 491 bp of exons (exons 1 to 3). In the 32 chickens (a total of 64 gene segments), we identified 19 SNP or 1 SNP per 200 bp (Figure 3.1). Deletions or insertions were absent. Ten of the 19 SNP were in the promoter region, 8 in intron regions, and only 1 in exon regions (exon 2, position 253). When corrected for the number of bases analyzed, the frequencies of SNPs in the promoter and intronic regions were about the same, whereas the frequency in exon region was more than two-fold lower. The SNP in exon 2 was a synonymous mutation, resulting in a change from an AAA triplet to an AAG, both coding for lysine. Variations in the promoter region were outside consensus sequences, a possible exception being a mutation 10 bases downstream of the TATA box. As expected, the majority of mutations were transitions. Nine of the SNPs were purine transitions, 6 were pyrimidine transitions and 4 were transversions.

To resolve double heterozygotes, it was assumed that each mutation occurred only once during evolution. It entails that only three of the four possible combinations of alleles at two loci are present. These three combinations were inferred from chickens where one of the two loci was homozygous. This assignment was consistent and delineated six different haplotypes labeled A1 to A6 respectively (Table 3.1). The sequence published by Sato *et al.* (1997) was not identical to any of the haplotypes we found in this research, but was closest to A3, differing by 4 base pairs. There was at least one homozygous chicken for each haplotype in our samples.

The most parsimonious unrooted gene tree is shown in Figure 3.2. It was established by inspection and coincided with the gene tree generated by Clustal W version 1.81[4]. At each polymorphic site the ancestral base was the most frequent base

among the 6 different haplotypes (Table 3.1, Figure 3.2). The same observation has been made in human and indicated that on an evolutionary scale these polymorphisms have occurred relatively recently (Cargill *et al.*, 1999). The 3' and 5' nearest neighbor bases were analyzed for their effect on the mutation frequencies as described by Hacia *et al.* (1999). Similar to the human study, we found that after correction for dinucleotide frequencies, the mutation rate at CpG dinucleotides was almost 10-fold higher than at other dinucleotides.

The C to T transition at locus R41 corresponds with a loss of an *Aci*I restriction site (AACGTT→ AATGTT) and the G to A transition at locus F22 with an acquisition of a *Bst*EII restriction site (GGTGGCC→ GGTGACC). These RFLPs can be used for rapid screening by PCR and to group haplotypes. The RFLP haplotype {*Aci*I- , *Bst*EII-} was diagnostic for A1, {*Aci*I+ , *Bst*EII-} for either A2, A3 or A4 and {*Aci*I+ , *Bst*EII+} for A6 or A5 (Figure 3.2). We have determined the two RFLP genotypes in 475 chickens of a non-selected random bred strain of White Leghorn chickens (Gowe *et al.*, 1993). As expected from the assumption that each mutation occurred only once in evolution, the genotypes were consistent with the presence of the three RFLP haplotypes indicated above, whereas the haplotype {*Aci*I- , *Bst*EII+} was absent. Linkage disequilibrium analysis indicates that this conclusion is statistically significant (data not shown).

We are now using this information to develop strains with the most distant haplotypes of PEPCK-C in order to test whether the genetic variability in this gene contributes to phenotypic variations. We are particularly interested in establishing whether there is epistasis with PEPCK-M the mitochondrial form of PEPCK. PEPCK-M

segregates for alleles which are co-selected with resistance to Marek's Disease (Li *et al.*, 1998a).

3.5 ACKNOWLEDGMENTS

Research support from the National Science and Engineering Council of Canada, Agriculture and Agri-Food Canada and Shaver Poultry Breeding Farms Ltd. is gratefully acknowledged.

---

[2]DNAMAN, Lynnon BioSoft, Quebec, CA, J7V 8P5

[3] QIAGEN Inc., Ontario, CA, L5N 7X9

[4] Biology Work Bench, http://workbench.sdsc.edu.

TABLE 3.1. Sequence variation and haplotypes in chicken cytosolic phosphoenolpyruvate carboxykinase.

Site	Position[1]	Haplotypes						
		A1	A2	A3	A4	A5	A6	X^2
F11	-1659	G	G	G	G	A	A	G
F12	-1595	A	A	A	A	G	G	A
F13	-1276	A	A	G	A	A	A	A
F14	-1146	C	C	G	C	C	C	C
F15	-1128	C	C	T	T	T	T	T
F21	-874	T	T	C	C	C	C	C
F22	-664	G	G	G	G	A	A	G
F23	-575	A	A	G	G	G	G	G
F24	-498	G	G	G	G	T	T	G
R21	-46	A	A	A	G	A	A	A
F31	242	C	C	C	C	G	G	C
F32	253	G	G	A	A	A	A	A
F33	352	G	G	G	G	A	G	G
R33	755	T	T	T	T	C	C	T
R32	863	T	T	T	T	C	C	T
R31	881	A	A	G	G	G	G	G
F41	1412	A	A	T	A	A	A	A
F42	1621	C	C	T	T	T	T	T
R41	1808	T	C	C	C	C	C	C

[1] The position numbering follows the sequence numbering of Sato *et al.* (1997). Transcription starts at position 1.

[2] Putative branch point.

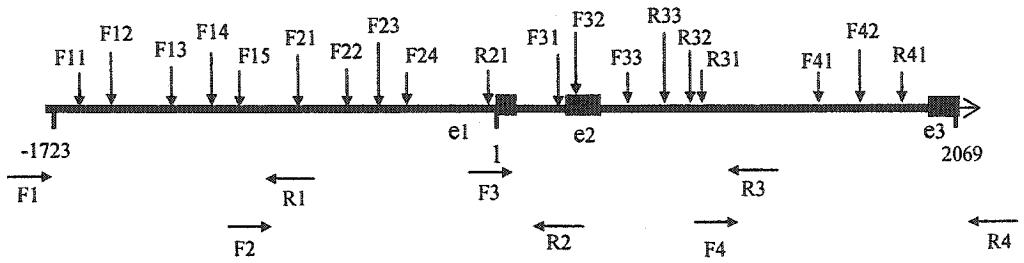


FIGURE 3.1. Sequence strategy and location of DNA polymorphic sites in chicken cytosolic phosphoenolpyruvate carboxykinase. The upper arrows show the locations and names of the polymorphic sites and the lower arrows the locations and names of the primers. The point of the arrow indicates the 3'-end of the primer. The lengths of the arrows are not to scale. e represents exon. The primer sequences were:

F1; CTGGGACCACCAGCAAGTACTG, **R1**; GCCTGTGCAGTCGGTGTGTGA,

F2; GCTGGGACTGAATGGAAGAGGAG, **R2**; CTGTTGAGTCGGATGGGTGTCAG,

F3; CACCATCAGCTGAAAGGGAGCC, **R3**; GTTGGGTTCGTTGGGAGAGACAAC,

F4; GTCTCTCCCAACGAACCCAACATG, **R4**; CCTCTTCTGACATCCAGCGACC.

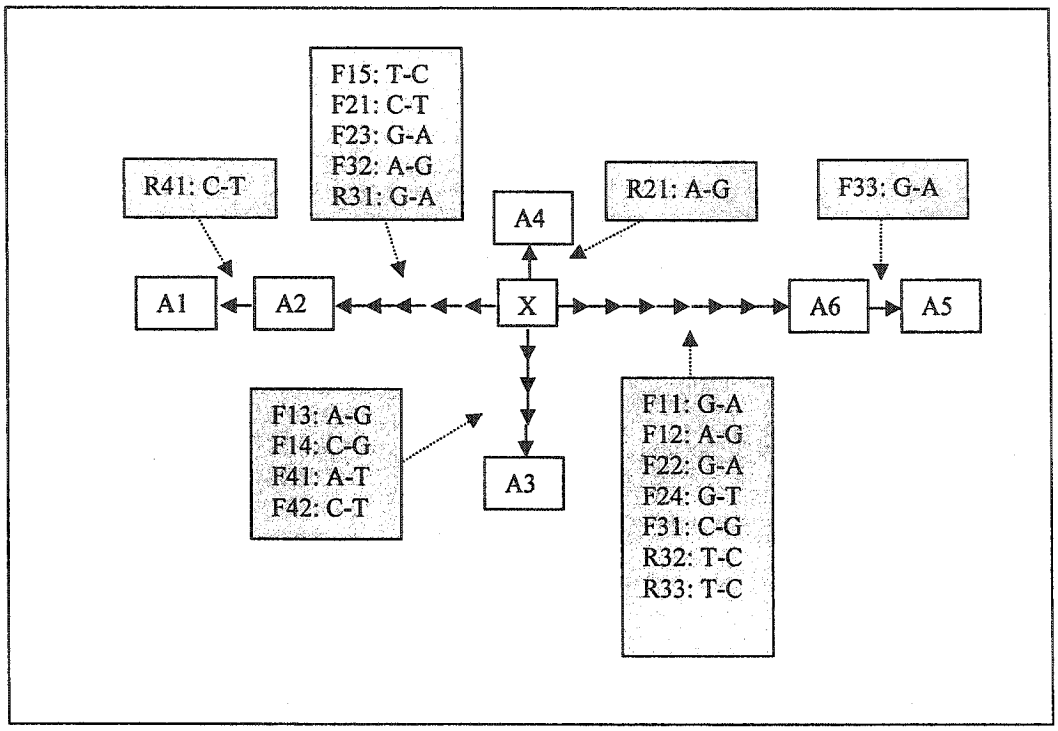


FIGURE 3.2. Evolutionary relationship between chicken cytosolic phosphoenolpyruvate carboxykinase haplotypes. The difference between two most distant haplotypes (A1 and A5) is 14 single nucleotide polymorphisms. X is a putative branch-point haplotype.

CONNECTING STATEMENT 1

Chapter 4 is a manuscript that has been published in Poultry Science (Parsanejad et al., 2003). Citations in this chapter are listed in the references at the end of the thesis. All tables and figures are presented at the end of the chapter.

In chapter 3, we described the identification of 19 single nucleotide polymorphisms (SNP) and 6 haplotypes in the 5' region of the PEPCK-C gene. Two SNPs coincided with RFLP at an *Aci*I and *BstE*II respectively. They were used to subdivide the PEPCK-C alleles into three haplotype classes. The next chapter describes the trait association of the genotypes defined by these haplotypes and their interaction with a marker of PEPCK-M, the mitochondrial form of PEPCK.

CHAPTER 4

ALLELES OF CYTOSOLIC PEPCK: TRAIT ASSOCIATION AND INTERACTION WITH MITOCHONDRIAL PEPCK IN A STRAIN OF WHITE LEGHORN CHICKENS

R. Parsanejad, A. Torkamanzehi[1], D. Zadworny and U. Kuhnlein[2]

Department of Animal Science, McGill University,

Ste. Anne de Bellevue, QC, Canada H9X 3V9

[1]Current address: Faculty of Agriculture, University of Sistan and Beluchistan, Zahedan, Iran.

[2]To whom correspondence should be addressed: kuhnleinu@macdonald.mcgill.ca

Published by Poultry Science (2003. 82(11):1708-15).

4.1. ABSTRACT

White Leghorn chickens from a non-selected closed population were typed for two RFLP located in the 5' end of the gene coding for cytosolic PEPCK (PEPCK-C), a major control gene of gluconeogenesis. The two RFLP gave rise to three alleles (or haplotype classes), which in turn defined six genotypes. A comparison of genotypes revealed significant differences in feed efficiency (FE) and residual feed consumption (RFC). FE is the ratio between feed intake and egg mass produced, while RFC is the feed intake after correcting for body weight and egg production. There was significant interaction between PEPCK-C genotypes and mitochondrial PEPCK (PEPCK-M) genotypes defined by an RFLP. The latter enzyme catalyzes the same reaction, but is located in the matrix of the mitochondria and is encoded by a different nuclear gene. Interaction was evident from an analysis of the egg weight and egg-specific gravity in the early phase of egg laying. It appeared to be non-additive, i.e. whether the genetic variation in one gene affected a trait or not depended entirely on the genotype of the second gene. In addition, significant genetic disequilibria were observed for two of the three alleles of the cytosolic PEPCK and between one of these alleles and the mitochondrial PEPCK RFLP. It indicates variations of genes in the gluconeogenesis pathway may affect feed utilization and egg production traits, as well as reproductive fitness.

4.2 INTRODUCTION

Phenotypes are determined by networks of interacting metabolites, proteins, RNA and DNA. Such interactions occur between and within all levels. Enzymes regulate metabolic processes and are themselves subject to regulation by metabolites. Proteins including enzymes control their own expression by regulating processes such as transcription, translation, RNA splicing and RNA editing. At a higher level, phenotypes determine the reproductive success of an individual in a breeding population. Hence they determine the repertoire of genetic variations present in a breeding population. The stability of such networks is provided by a myriad of regulatory loops and feedback controls.

Due to the complexity of the relation between genotype and phenotype, the search for quantitative trait mutations has been shifted from segregation analyses to association studies. Association studies entail the matching of genetic patterns with phenotypic patterns. Such patterns, however, are difficult to characterize. At the phenotypic level, patterns may be classified by techniques such as cluster analysis or principal component analysis (Dopazo *et al.*, 2001). At the genetic level, nucleotide variations may be grouped into haplotypes (i.e. arrays of alleles that are co-inherited). In addition, one may pinpoint sites of recombination and further refine of the genotypic profile. The major problem is to create a classification of phenotypes and genotypes that are meaningful, yet broad enough to contain a sufficient number of observations for statistical analysis.

We have previously analyzed White Leghorn chickens for DNA polymorphisms in the gene coding for the cytosolic form of phoshoenolpyruvate carboxykinase (GTP) [EC 4.1.1.32] (PEPCK) (Parsanejad *et al.*, 2002). PEPCK is a key regulatory enzyme of gluconeogenesis, catalyzing the formation of phosphoenolpyruvate by decarboxylation of

oxaloactetate while hydrolyzing GTP (Hanson and Reshef, 1997). There are two isozymes encoded by two different nuclear genes. One isozyme (PEPCK-C) is located in the cytosol (Cook *et al.*, 1986), and the other (PEPCK-M) in the mitochondrial matrix (Weldon *et al.*, 1990).

Sequence analysis of a 4 kb fragment of PEPCK-C in White Leghorns revealed the presence of 19 single nucleotide polymorphisms (SNP) that could be grouped into 6 different haplotypes (combinations of SNP alleles). Two of the SNP coincided with RFLP, which enabled us to further subdivide the 6 haplotype classes into three classes, representing different branches of the gene tree (Parsanejad *et al.*, 2002)

Here we investigated the association of the 6 PEPCK-C genotypes defined by two RFLP with production traits in a strain of White Leghorn chickens. We were particularly interested to investigate whether this analysis coupled with the knowledge of the gene tree would enable us to identify the branches of the gene tree that harbored the mutation responsible for trait differences.

We also report on the interaction of PEPCK-C with the isozyme PEPCK-M. The latter gene was subdivided into genotypic classes on the basis of a single RFLP. There was significant genotypic disequilibrium between the PEPCK-M alleles and one of the three alleles of PEPCK-C, indicating reduced fitness for some allele combinations. Interaction was also evident from a comparison of trait values in the genotypic classes formed by the two loci.

4.3. MATERIALS AND METHODS

4.3.1 Experimental chickens and traits

Chickens were from a White Leghorn strain (strain 7) which had been established from North American commercial stocks in 1958 (Gowe & Fairfull, 1990; Gowe *et al.*, 1993). It consisted of 100 sire families and was propagated by mating one sire with two randomly chosen females. The chickens were weighed prior to the onset of egg laying (140 d of age, HBWT) and at 265 d of age (MBWT). The age of the onset-of laying (AFE) was recorded. Egg weight (EWT) and specific gravity (SPG) were measured on eggs laid during five consecutive days over three time periods starting from 240d (HD1), 350d (HD2) and 450d (HD3) and averaged for each period. The rate of egg laying was measured between AFE and day 275, 276d-356d and 357d-457d. Feed consumption and egg mass (EGM) were measured between 247 days and 268 days of age. Feed conversion (FC) is the feed intake divided by the egg mass. The residual feed consumption (RFC) is the feed intake after correcting for body weights and egg mass produced. In our case we conducted a regression analysis of the feed intake on the MBWT and EGM for the most frequent PEPCK genotype class (class A/C). These regression parameters were used to calculate the expected feed consumption for each individual. RFC is the difference between the actual feed consumption and the expected feed intake.

4.3.2 Detection of RFLP

The primers and PCR conditions are given in Table 4.1.

4.3.3 Statistical analysis

For statistical analyses and graphics we used the NCSS program (Hintze, 1997). Unless otherwise stated we used rank based statistics to avoid the necessity to specify biologically unrealistic models and to avoid problems with distributions that deviated from normality. However, parametric tests (GLM procedures, ANOVA or t-tests) gave the same result, even when traits deviated from normality or subgroups had significantly different variances. Four outliers with AFE>200 days and one with HD1< 50% were removed from the data set.

Composite genotypic disequilibria between the three alleles of PEPCK-C and the two alleles of PEPCK-M were analyzed using the program of Lewis and Zaykin (2001). In this program, the expected frequencies of each allele are determined and used to calculate the expected allelic combinations. The disequilibrium coefficients are the deviations of the expected from the observed frequencies of allele combinations. We used the program that screens pairs of alleles, one at each locus. The disequilibrium is called genotypic, since it does not assume Hardy-Weinberg equilibrium and composite, since the two types of double heterozygote are lumped together. The program yields 6 disequilibrium coefficients. For two loci 1 and 2, each with a '+' allele and a '–' allele, the two coefficients $D_{1(+)}$ and $D_{2(+)}$ are the deviations from Hardy-Weinberg equilibrium at each locus, $D_{1(+)2(+)}$ the deviation of the digenic combination of the '+' alleles at the two loci, $D_{1(+)2(+)2(+)}$ and $D_{1(+)1(+)2(+)}$ the two possible combinations of the three '+' alleles and $D_{1(+)1(+)2(+)2(+)}$ the combination of all four '+' alleles. Each disequilibrium coefficient is corrected by the preceding lower disequilibrium. The program lists the disequilibrium coefficients for the combinations of the '+' allele. Their relationship to coefficients of mixtures of '+' alleles and '-' alleles is easily deduced. In our case we will

be interested in the trigenic disequilibrium $D_{1(+)\,2(+)\,2(+)}$. It can be shown that $D_{1(+)\,2(+)\,2(+)} = -D_{1(+)\,2(+)\,2(-)} = D_{1(+)\,2(-)\,2(-)} = -D_{1(-)\,2(+)\,2(+)} = D_{1(-)\,2(+)\,2(-)} = -D_{1(-)\,2(-)\,2(-)}$. Hence if the gene combination 1(+) 2(+) 2(+) are in excess, the combinations 1(+) 2(-) 2(-) and 1(-)2(+)2(-) are also in excess, while all the other gene combinations are at a deficit.

4.4. RESULTS AND DISCUSSION

4.4.1. Frequencies and trait association of PEPCK-C alleles

A total of 395 hens from a White Leghorn strain (strain 7) were typed for the RFLP at the *Aci*I site and the RFLP at the *Bst*EII site (Parsanejad *et al.*, 2002). The absence or presence of restriction sites at the two loci defined three haplotype classes: haplotype A [*AciI*$^+$ *Bst*EII$^+$] , haplotype B [*AciI*$^+$ *Bst*EII$^-$] and haplotype C [*AciI*$^-$ *Bst*EII$^-$] (Table 4.2). The fourth combination [*AciI*$^-$ *Bst*EII$^+$] was absent as expected from parsimony during evolution and the lack of recombination between alleles. Based on a previous analysis of the mutation spectrum, haplotypes A and C represent two different branches of the gene tree and were genetically most distant. Haplotype B comprised haplotypes closer to the putative ancestral PEPCK-C gene, [A ← B → C]. For convenience we will subsequently refer to the three haplotype classes as alleles A, B and C. The frequency of alleles as well as of the genotypes did not differ significantly between males and females.

The six different genotypes arising from the three alleles were analyzed for association with production traits (Table 4.3). Among these were the body weight, the rate of egg laying, egg weight and egg-specific gravity. They constitute the major outlays of metabolic energy in a laying chicken. Specific gravity is an indirect measure of the egg-shell calcium, which in chickens amounts to a daily output of about 10% of the total body calcium (Wasserman & Fullmer, 1989). Other traits analyzed were the age of sexual maturity (age at first egg), the feed intake, the feed efficiency (ratio between feed consumption and egg mass produced) and the residual feed consumption (RFC). The

residual feed consumption is the difference between the actual feed intake and the expected feed intake after correcting for body weight and egg mass produced (Morisson *et al.*, 1997).

Non parametric tests revealed significant differences in the median values for the feed efficiency and the residual feed consumption, but not for any of the other traits. Contrasts between genotypic classes indicate that none of the three alleles of the PEPCK-C genotypes were equivalent. Specifically, the median for RFC was lower for B homozygotes than A and C homozygotes, indicating that the more ancient allele B is different from A and C. The median also differed between AB and BC heterozygotes indicating that alleles A and C also differ (Figure 4.1).

The median is a measure of the central tendency of the distribution. In order to test whether other aspects of the distributions of RFC were affected we compared quantile plots of the trait distributions (Figure 4.2). The distributions for the genotypes AC, BC and CC were nearly identical, indicating that allele C has a dominant effect on the trait distribution. The distributions for the genotype BB and AB and BC differed significantly, indicating again that alleles A, B and C in conjunction with allele B are not equivalent.

The distribution of body weight and egg mass, the independent variables used to determine the predicted feed intake, did not vary among the different genotypic classes. However, there were some differences in feed intake. The AB chickens which had a higher RFC tended to have higher feed intake, while the BB chickens which had a lower RFC tended to have a lower feed consumption (Figure 4.2).

The distribution of RFC values indicates how well the data of the different genotypic classes fits the regression equation of feed consumption on egg mass and body weight derived from the entire strain (or as here one of the genotypic classes). It indirectly

tests for differences in the elevation and the slopes of the regression equations derived separately for each of the genetic groups. A graphic presentation of the partial regression lines of the feed consumption on egg mass and body weight is shown in figure 4.3. The BB genotypes have the lowest elevations and the highest partial regression coefficients, for both egg mass and body weight. A lower elevation indicates that the feed required to maintain body weight and produce egg mass is lower, an increased slope indicates that, as egg mass or body weight increase, the conversion becomes less efficient. For egg mass, BB appears to be more efficient than the other genotypes throughout the entire range, while for body maintenance of body weight, BB is only more efficient for individuals with a low but not with the high body weight. The genotypic class AB, which also markedly differed in the distribution of RFC values, had a lower elevation and higher slope for egg mass. However, in this case the feed consumption was higher throughout the entire range of egg mass values. The partial regression on body weight was similar to the majority of genotypes.

4.4.2. Deviation from Hardy-Weinberg equilibrium

The PEPCK-C genotypes defined by the three alleles deviated from Hardy-Weinberg equilibrium (Hernandez and Weir, 1989). There was an excess of heterozygotes AC and a deficit of the homozygotes AA and CC (Table 4.4). In contrast, the heterozygotes AB and BC occurred at the expected frequencies. Deviations from equilibrium were not observed for other genes, such as GH, ODC, IGF-1 and PEPCK-M (Table 4.5). Since the strain was propagated by random mating at an effective population

size exceeding 500, we assume that the deviation from equilibrium is due to differential fitness.

A comparison between the genotypic class AC (excess) with the combined genotypic classes AA and CC (deficient) revealed similar medians for the traits listed in Table 4.3, with the exception of the housing body weight (t-test, P=0.024). A normality plot of the distributions is shown in figure 4.4. It revealed that above the median the distribution among the AC genotypes is shifted to lower values, while below the median the two distributions are similar. Hence, PEPCK-C may be rate limiting in chickens with high, but not with low body weight.

It is unlikely that the reduction in housing body weight is directly associated with reduced reproductive fitness. Rather, PEPCK-C in conjunction with other genes may affect sperm count, fertility or hatchability via other pathways. Divergent selection for residual feed consumption, the major trait affected by PEPCK-C, has been reported to reduce reproduction and alter sperm characteristics (Morisson *et al.*, 1997).

4.4.3. Interaction with the mitochondrial form of PEPCK (PEPCK-M)

The mitochondrial form of PEPCK carries out the same reaction as the cellular form of PEPCK but is encoded by a different nuclear gene. In order to test for interaction between the two genes, we made use of an *Acc*I RFLP marker in PEPCK-M gene to subdivide the PEPCK-M genotype into three classes, the two classes homozygous for opposite alleles and the heterozygotes. These genotypic classes by themselves were not associated with differences in the medians or means of the production traits listed in Table 4.3 (data not shown). The three PEPCK-M genotypes, together with the six

PEPCK-C genotypes give rise to 18 different combinations. To reduce the number of combinations and increase the number of observations in each class, we analyzed the combination of the genotypes for one PEPCK-C allele at a time ([A] vs. [B,C]; [B] vs. [A,C]; and [C] vs. [A,B]) This yields 9 combinations of PEPCK-C/PEPCK-M genotypes for each PEPCK-C allele. In addition, genotype combinations where the number of observations was less than 5% were omitted from the analysis (Table 4.6).

For the PEPCK-C allele A, all the genotype combinations had similar median values for the production traits listed in Table 4.3. For the allele B a significant difference was observed for the egg weight measured at 240 days of age. Analysis of contrasts indicated that the PEPCK-M genotype affected the egg weight among the PEPCK-C genotypes B+/- but not B-/-. The genotypes B+/+ were too rare for a meaningful analysis.

Analyses using of the genotypic combinations of PEPCK-M with genotypes of PEPCK-C defined by the C-allele indicated a significant association with egg specific gravity. The PEPCK-C genotypes C+/+, C+/- and C-/- only affected egg specific gravity in combination with the PEPCK-M genotype *Acc*I +/+ and *Acc*I -/-.

4.4.4. Linkage disequilibrium between PEPCK-C and PEPCK-M

Similar to the deviation from Hardy-Weinberg equilibrium that we observed for PEPCK-C alleles we would expect interaction of that gene with PEPCK-M leads to genotypic disequilibria. Genotypic disequilibria refer to the deviation of the observed from expected allelic combinations. In a two locus / two allele system genotypic disequilibria can be characterized by six parameters, two single gene disequilibrium coefficients (Hardy-Weinberg disequilibrium), a composite digenic disequilibrium coefficient, two trigenic coefficients and a composite quadrigenic disequilibrium

coefficient. In our case, genetic variation at PEPCK-C was characterized by three alleles, giving rise to three different sets of parameters.

The significance (chi-square values) of the parameters for the PEPCK-M / PEPCK-C interaction are given in Table 4.7. The single genotypic disequilibria were significant for PEPCK-C homozygotes A +/+ and C +/+ and indicate a deficiency as already indicated in Table 4.4. Among the higher order disequilibria, the trigenic disequilibrium D_{122} was significantly positive (P=0.011) for the PEPCK-C allele C, indicating an excess of the gene combination [PEPCK-M: *AccI*+ ; PEPCK-C: C+/C+].

The disequilibria listed in Table 4.7 are for the + alleles. They are inter-related with disequilibria of mixtures of + and – alleles (see Materials and Methods). Among the nine genotype combinations of the C-alleles of PEPCK-C and *AccI* alleles of PEPCK-M (see Table 4.6), three contain only the tri-allelic combinations which are in excess and three the tri-allelic combinations which are deficient. Genotypes with allelic combinations at an excess are [PEPCK-M *AccI*+/+ PEPCK-C *C*+/+], [PEPCK-M *AccI*+/+, PEPCK-C *C*-/-] and [PEPCK-M *AccI* -/- PEPCK-C *C*+/-] and genotypes with allelic combinations at a deficit are [PEPCK-M *AccI*+/+ PEPCK-C *C*+/-], [PEPCK-M *AccI*-/-, PEPCK-C *C*-/-] and [PEPCK-M *AccI* -/- PEPCK-C *C*+/+]. A comparison of the pools of these two classes revealed that genotypes with allelic combinations in excess had a higher egg specific gravity in period 1 than those with allelic combinations in deficit. The median values were 86 and 84 units, respectively (P=0.009). No differences were observed in period 2 and period 3.

4.5. CONCLUSION

We have shown that PEPCK-C, a key regulatory gene in gluconeogenesis affects RFC. Residual feed consumption is a measure of how efficient feed is utilized to maintain body weight and produce eggs. In the strain analyzed, two genotypes of PEPCK-C, AB and BB had different medians and distributions of RFC than the other genotype combinations. Both, the genotypes did not affect the distribution of egg production or body weight, but there was a slight (but not significant) effect on feed intake that paralleled the differences of in RFC values. It may therefore be the genetic effects on RFC values are at the level of feed digestion. However, regression analyses also point towards differences in partitioning of feed energy into different metabolic processes. In particular, partial regression of feed consumption on body weight and egg mass indicate that PEPCK-C genotypes have different effects on slopes as well as elevations.

Since PEPCK-C variants were associated with changes in a very basic metabolic parameter, that there was little effect on other production traits. It indicates that such traits are well buffered against variations in energy metabolism. Some of this buffering may be mediated by genetic variants in other genes. We have analyzed the interaction of PEPCK-C with a marker in PEPCK-M. It indeed showed significant effects on egg weight and egg specific gravity for some marker combinations (Figure 4.5). Although we have singled out PEPCK-M because it is an isozyme of PEPCK-C, we have observed similar effects with other genes (unpublished results). It supports the notion that networks involving many genes stabilize complex traits.

The three RFLP, two in PEPCK-C and one in PEPCK-M, were chosen because they provide a convenient assay to genotype a large number of individuals. They are a random sample of SNP present in these genes. We were therefore surprised that each

genotypic class defined by these markers had different genotypic properties. It therefore seems that the number of variants that affect phenotypes is large, even within a single gene. The multiplicity of alleles and consequently of genotypes is a formidable problem in association studies, since it increases the number of different groups while decreasing the number of observations in each group.

Epistatic processes may be the active principle maintain some of these variants, leading to what has been called genetic homeostasis. Indeed, genetic disequilibria indicate that some allele combinations are present at a higher frequency than expected from random association of gametes. The disequilibrium between PEPCK-C and PEPCK-M described here is not an exception but can be observed with many other genes (unpublished results).

Disequilibrium analysis may be used as a rational to group genotype combinations for association analysis. Specifically, genotypes with allele combinations that are in excess can be compared with genotypes with allele combinations which are at a deficit. The rational is that most gene combinations that have major effects will also have an impact on physiological processes that affect fitness. In our example, comparison of the PEPCK-C A/C heterozygote (excess) with the corresponding homozygotes (deficit) indicated differences in the housing body weight. Similar pooling on the basis of the trigenic disequilibrium between PEPCK-C and PEPCK-M indicated differences in egg specific gravity (Figure 4.6). We have extended this method to other gene combinations and so far found trait associations in every single case (unpublished results).

4.6 ACKNOWLEDGMENTS

Research support from the National Science and Engineering Research Council of Canada, Agriculture and Agri-Food Canada and Shaver Poultry Breeding Farms Ltd. is gratefully acknowledged.

TABLE 4.1. PCR primers and conditions[1] for cytosolic and mitochondrial forms of phosphoenolpyruvate carboxykinase.

RFLP[2]	Primer	Sequence
PEPCK-C *Aci*I	Forward	GTCTCTCCCAACGAACCCAACATG
(pos. 1808)	Reverse	CCTCTTCTGACATCCAGCGACC
PEPCK-C *BstE*II	Forward	GCTGGGACTGAATGGAAGAGGAG
(pos. -664)	Reverse	CTGTTGAGTCGGATGGGTGTCAG
PEPCK-M *Acc*I	Forward	CCTTCGCCATGAGCCCCTTTTTC
(pos. 1578)	Reverse	CAGCTCCGCCATGACATCCCT

[1] PCR conditions: 5 min presoaking at 95° followed by 35 cycles of 60 sec at 94°; 80 sec at 62° (PEPCK-C) or 60° (PEPCK-M), and 90 sec at 72°.

[2] The PEPCK-C *Aci*I and the PEPCK-C *BstE*II RFLP were located at position 1808 and -664, respectively (Parsanejad *et al.*, 2002). The PEPCK-M RFLP was located at position 1578 (Weldon *et al.*, 1990).

TABLE 4.2. Definition and frequency of PEPCK-C haplotypes

Haplotype	RFLP-alleles		Haplotype Frequencies	
	*Aci*I	*BstE*II	Females (N=350)	Males (N=119)
A	+	+	0.290	0.286
B	+	-	0.164	0.214
C	-	-	0.546	0.500

TABLE 4.3. Comparison of the trait medians of the PEPCK-C genotype classes[1]

Trait	PEPCK-C genotype[2]						Significance (P-value)
	AA 21-19	AB 29-23	AC 125-114	BB 10-7	BC 62-48	CC 88-74	
Body Weight(g) at							
130 d	1310	1265	1270	1220	1310	1305	0.195
265 d	1781	1772	1745	1700	1764	1777	0.150
Age at First Egg (d)	169.0	168.0	167.0	168.0	163.5	165.5	0.539
Egg Weight (g)							
Per.1	52.6	51.4	52.3	53.0	52.7	52.4	0.698
Per.2	57.2	56.6	57.7	59.4	58.2	58.2	0.593
Per.3	61.5	61.0	60.0	62.5	59.8	61.2	0.495
Egg Specific Gravity							
Per.1	86.0	86.0	84.7	86.8	85.0	85.2	0.268
Per.2	82.0	82.0	82.0	80.7	82.0	82.0	0.728
Per.3	80.0	78.0	78.5	76.0	79.3	80.0	0.608
Rate of egg laying (%)							
Per.1	86.7	84.1	84.2	89.1	84.2	84.6	0.285
Per.2	71.3	72.5	72.5	71.2	70.0	71.3	0.839
Per.3	65.0	60.0	60.0	50.6	55.0	57.5	0.369
Feed Consumption (g)	117	117	113	111	111	116	0.247
Feed Efficiency[3]	2.48	2.56	2.44	2.27	2.50	2.42	**0.026**
RFC (g)[4]	22.7	22.5	18.0	12.5	18.4	16.8	**0.028**

[1] The significance was tested using the Kruskal-Wallis analysis of variance by rank, since some traits were not normally distributed. Parametric analysis of the traits whose values were close to normally distributed gave the same results.

[2] The number of observations in each class are indicated below the genotype. Variations within a genotypic class are due to incomplete observations.

[3] Significant contrasts were AB>AC, BB, CC; AC>BB and BB<BC, CC (Kruskal-Wallis multiple comparison test).

[4]Deviation from the expected value calculated from the regression equation of the feed consumption on mature body weight and egg mass produced in the genotype class AC. Note, the value in class AC equals the 0-intercept in the regression equation. Significant contrasts were AA>BB; AB>AC, BB, BC, CC and AC>BB.

TABLE 4.4. Genotypes frequencies of PEPCK-C and deviation from Hardy Weinberg-equilibrium

Genotype	Observed Frequency	Deviation from expected frequencies	Significance (chi-square, df=1)[1]
A/A	0.060	- 0.024	9.13**
B/B	0.032	+ 0.004	0.67
C/C	0.266	- 0.030	4.74*
A/B	0.086	- 0.010	0.57
A/C	0.375	+ 0.059	7.74**
B/C	0.181	+ 0.001	0.01

[1]Determined as described by Hernandez and Weir (1989). ** $P<0.01$; *$P<0.05$.

Table 4.5. Deviation from Hardy-Weinberg equilibrium at different loci in females of strain 7

Locus	Marker[a]	D[b]	P-value[c]
PEPCK-C	*Aci*I	-0.030	**0.023**
	*Bst*EII	-0.023	**0.033**
PEPCK-M	*Acc*I	-0.014	0.304
ODC	*Hind*III	+0.020	0.103
	*Msp*I	+0.010	0.351
IGF1	*Pst*I	+0.010	0.274
GH	*Sac*I	+0.005	0.489
	*Msp*I	-0.005	0.400
Pit1	*SSCP-A*	+0.299	**$<10^{-6}$**
	SSCP-D	+0.107	0.717

[a] All markers are RFLP at the indicated restriction sites. Exceptions are two markers in *Pit1* which are single-strand conformational polymorphism (Kuhnlein *et al.* 2003)

[b] Deviation of the observed homozygotic frequency from the expected frequency calculated as the product of the gametic frequencies . A positive value indicates excess and a negative of homozygotes and a negative value a corresponding deficiency of heterozgotes.

[c] Fisher's exact test based on 3200 runs (Lewis & Zaykin, 2001)

TABLE 4.6. PEPCK-M and PEPCK-C interaction: Number of observations for each genotypic combination and association with production traits

PEPCK-C Genotype	PEPCK-M genotype						Trait with significant association[2]
	AccI +/+		*AccI*+/-		*AccI*-/-		
	N	Median[1]	N	Median	N	Median	
A +/+	10	-	9	-	2	-	None
A +/-	61	-	71	-	22	-	
A -/-	59	-	87	-	19	-	
B +/+	3	-	7	-	1	-	Egg weight per. 1 (P=0.017)
B +/-	30	50.5 g [a]	48	52.6 g [b]	14	52.4 g [b]	
B -/-	97	52.4 g [b]	112	52.8 g [b]	28	51.5 g	
C +/+	40	85.0	45	86.0 [b]	7	-	Egg spec. gravity per. 1 (P=0.022)
C +/-	63	83.3 [a]	93	85.2	29	86.0	
C -/-	27	87.0 *	29	85.6	6	-	

[1] Median value of traits which significantly varied among genotypes.

[2] Only the genotype combinations with >10 observations were included in the analysis. Significance was assessed using the Kruskal-Wallis One Way ANOVA on ranks. The traits analyzed were those listed in Table 2. Trait associations with P>0.1 were considered as not significant.

[a, b] Contrasts were analyzed for traits which showed significant association with genotype combinations at the two loci. Different superscripts indicate significance (Kruskal-Wallis Multiple-Comparison Z-value test).

* The combination PEPCK-C: C -/-, PEPCK-M: *AccI* +/+ had significantly higher median egg-specific gravity (period 1) than any other genotype combination.

TABLE 4.7. Genotypic disequilibrium between PEPCK-M and PEPCK-C.

PEPCK-C allele	Significance of linkage disequilibrium (chi-square values[1])					
	D_1[2]	D_2	D_{12}	D_{112}	D_{122}	D_{1122}
A+	-1.16	**-4.19**	+0.27	+1.41	+0.47	-0.43
B+	-1.16	+0.33	-1.79	-0.63	+0.14	-0.43
C+	-1.16	**-4.22**	+0.39	-0.19	**+6.41**[3]	-0.13

[1] The degree of freedom is one. The sign in front of chi-square indicates whether the observed frequency of an allele combination is smaller or larger than the expected frequency.

[2] The numeral subscript refers to the gene and locus. The subscript 1 stands for the *AccI*+ allele of the PEPCK-M gene and the subscript 2 for the PEPCK-C allele listed in the first column.

[3] A significant positive trigenic disequilibrium D_{122} means that the observed frequency of the allelic combination p_{122} (i.e. PEPCK-M *AccI*+, PEPCK-C C+/+) is higher than the expected frequency which is the product of the estimated allelic frequencies ($p_1 \cdot p_2 \cdot p_2$) corrected for the lower order disequilibria.

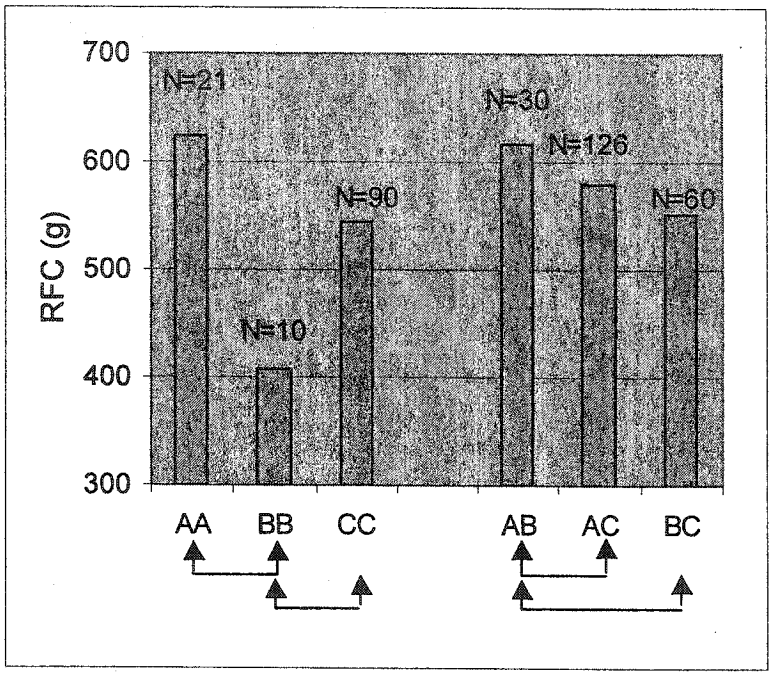


Figure 4.1. Comparison of the median RFC for different PEPCK-C genotypic classes. Arrows indicate contrasts that show that the alleles A, B and C are not equivalent. Other statistically significant contrasts are CC vs. AB and BB vs. AB and AC.

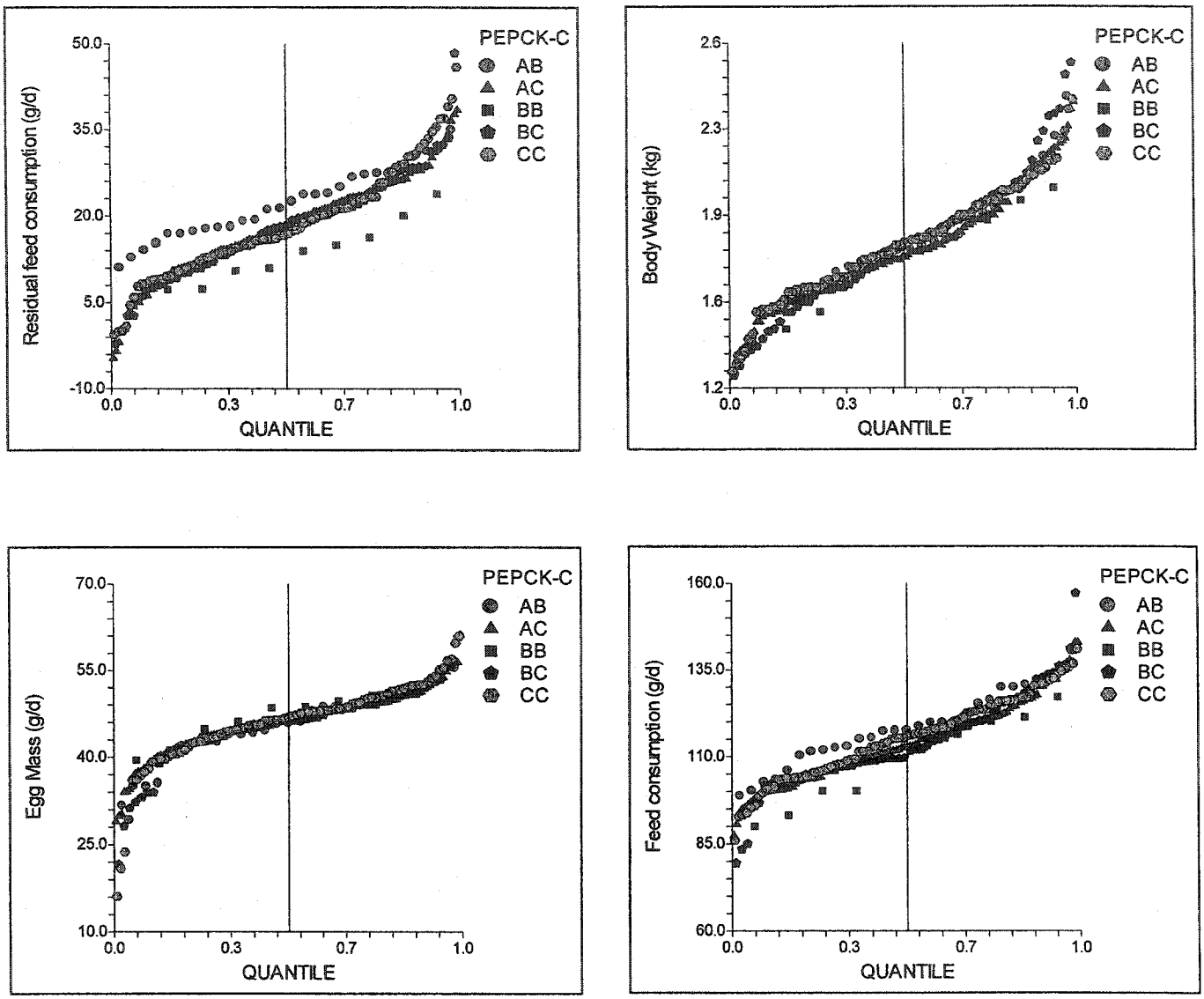


FIGURE 4.2. Quantile plots of the distributions of RFC, body weight at 260 d of age, egg mass produced and feed consumed. The distribution of the genotype AA is not shown for clarity. It follows the distribution of AC, BC and CC below the median and the distribution of AB above the median.

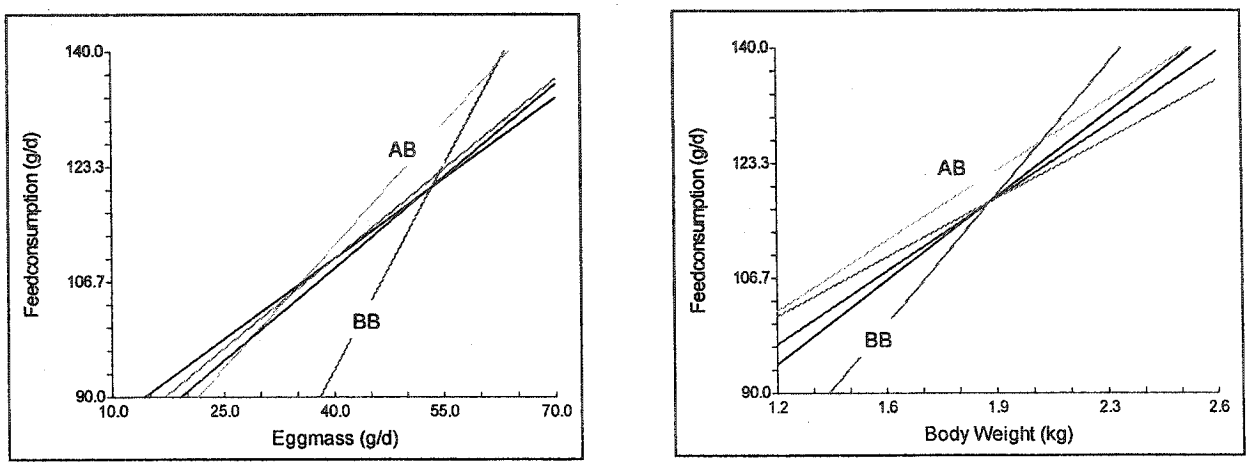


FIGURE 4.3. Partial regression lines describing the relationship between predicted feed consumption and egg mass and body weight for the genotypes AB, BB, AC, BC and CC. The predicted values obtained by regressing feed consumption on body weight and egg mass in each genotypic class were plotted against egg mass or body weight. Only the regression lines for the genotypes AB and BB are marked.

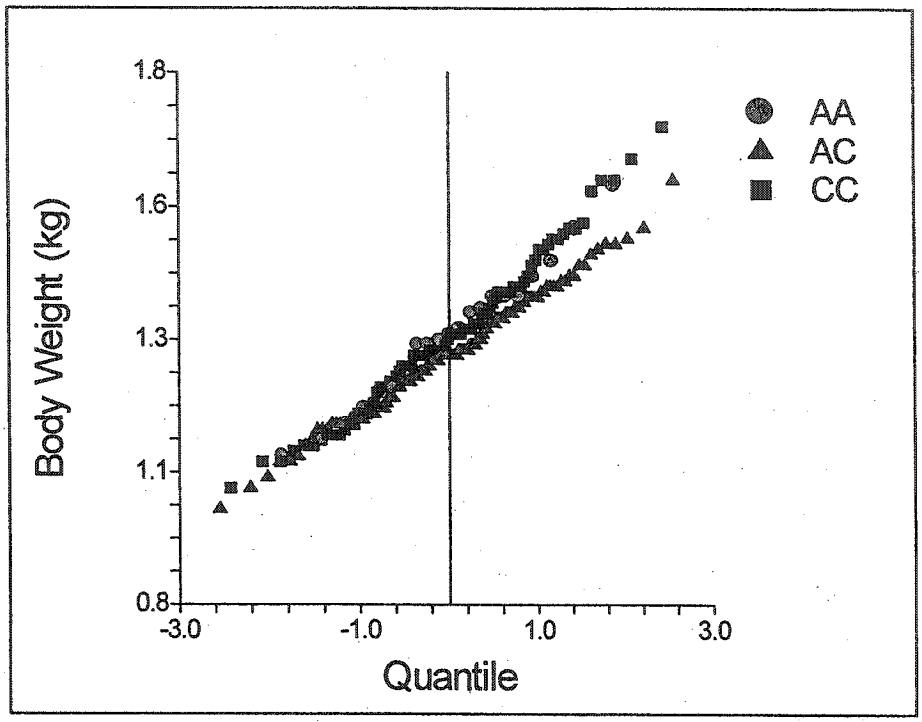


FIGURE 4.4. Normality plot of the juvenile body weight for the genotypic classes which deviate from Hardy-Weinberg disequilibrium. The intercept with the ordinate at 0 equals the median and the slope of the curve the standard deviation. The variances between AC vs. AA and CC combined differed at a P-level of 0.026 (Variance-Ratio Equal-Variance test) or 0.054 (Modified-Levene Equal-Variance test).

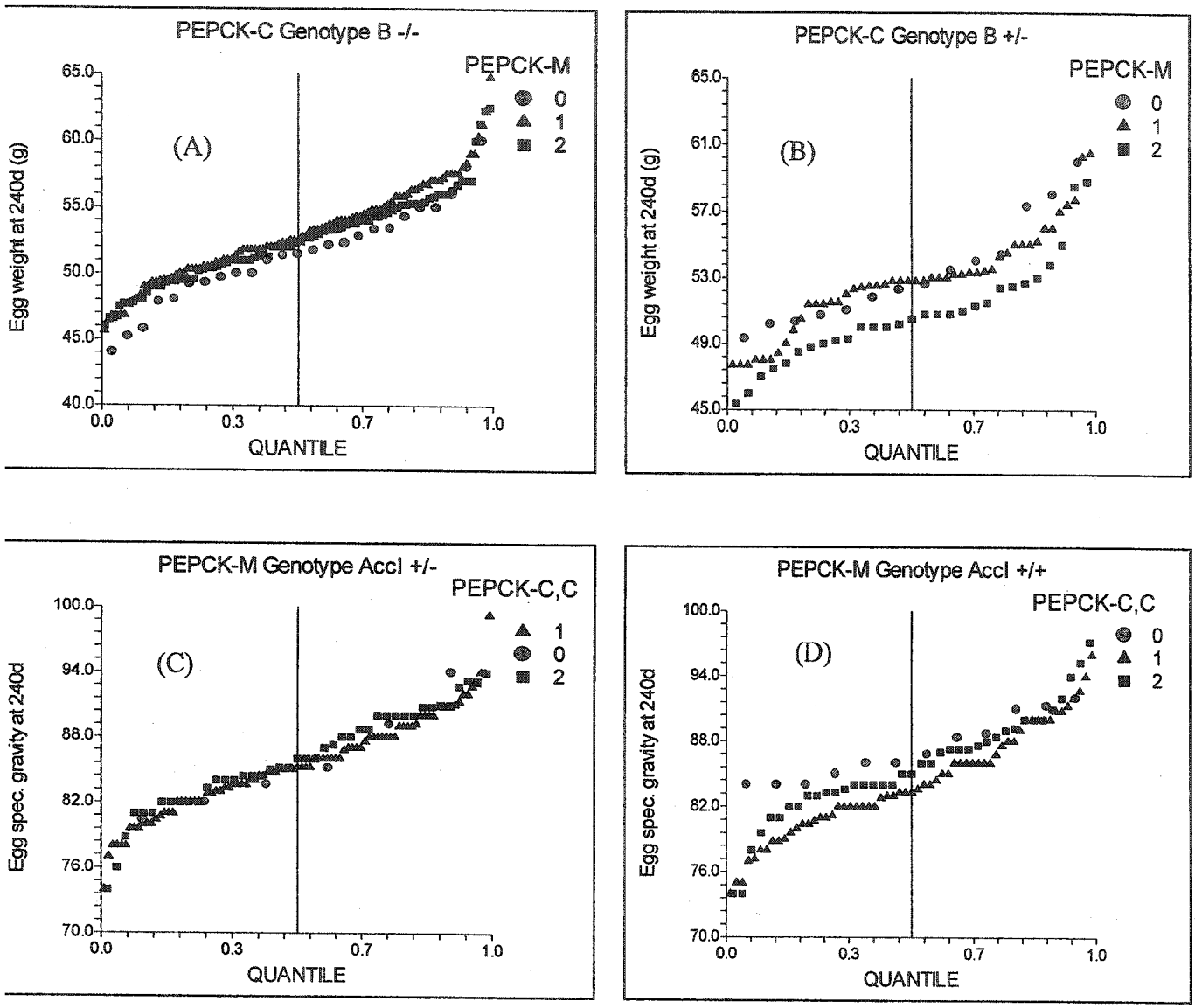


FIGURE 4.5. Interactive effects between PEPCK-C and PEPCK-M on egg weight and egg specific gravity. The upper panels show the egg weight distributions (240 d) for different PEPCK-M genotypic classes in two PEPCK-C allele B backgrounds (panel A: B+/-; panel B: B +/+). The lower panels show the egg specific gravity distributions (240 d) for genotypic classes for different PEPCK-C allele C genotypes in two PEPCK-M backgrounds (panel C: PEPCK-M *Acc*I +/-; panel D: PEPCK-M *Acc*I +/+). Genotypes -/- are designated as 0, genotypes +/- as 1 and genotypes +/+ as 2.

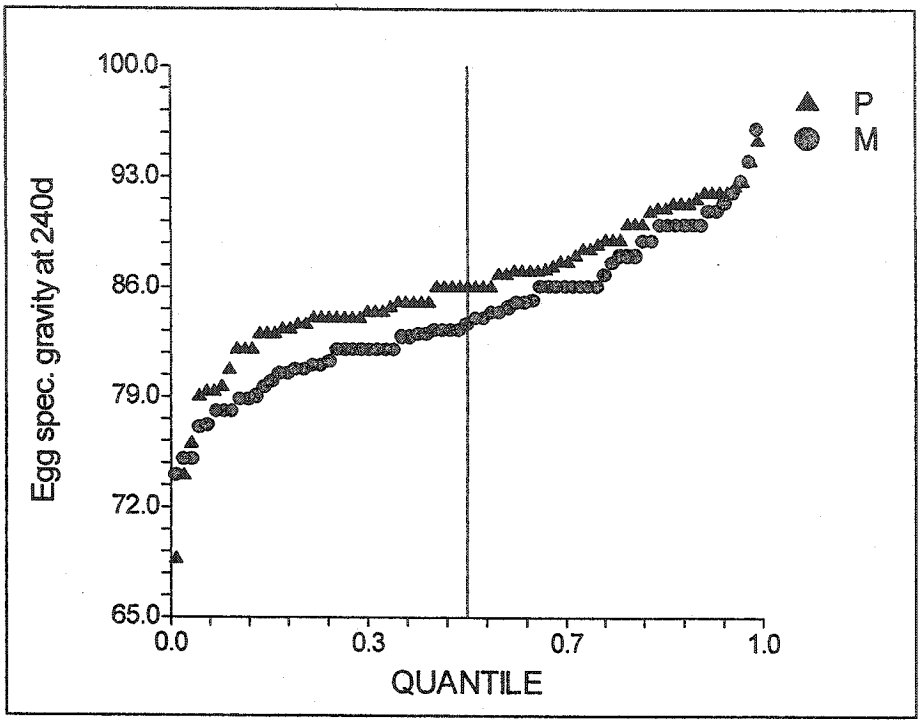


FIGURE 4.6. Distribution of specific gravity among genotypes grouped on the basis of linkage disequilibrium between PEPCK-C and PEPCK-M. (P) designates the pool of genotypes which contain only alleles which are at an excess and (M) genotypes with alleles which are at a deficit (M).

CONNECTING STATEMENT 2

Chapter 5 is a manuscript that has been accepted by Poultry Science. References are listed in the bibliography at the end of the thesis.

In chapter 3 and 4, we described the identification of DNA variants, haplotype establishment, detection of RFLP and their association with production traits for PEPCK-C. This chapter focuses on another candidate gene, ornithine decarboxylase (ODC), DNA variations and haplotype classification of the 3' end of ODC gene are reported. Two RFLP (*Msp*I and *Hind*III) diagnostic for the haplotypes were identified. They were used to analyze the association of ODC genotypes with production traits.

CHAPTER 5

ORNITHINE DECARBOXYLASE: HAPLOTYPE STRUCTURE AND TRAIT ASSOCIATION IN WHITE LEGHORN CHICKENS

R. Parsanejad, D. Zadworny and U. Kuhnlein[1]

Department of Animal Science, McGill University,

Ste. Anne de Bellevue, QC, Canada H9X 3V9

[1] To whom correspondence should be addressed: kuhnleinu@macdonald.mcgill.ca

Accepted by Poultry Science

5.1. ABSRACT

Ornithine decarboxylase (ODC), a rate-limiting enzyme in polyamine biosynthesis, has a significant role in DNA synthesis and cell proliferation. The genetic variability of the ODC gene, a candidate gene for cell growth, was analyzed in this communication. Five kb of the 3' end of the gene was sequenced in 20 individuals from a strain of White Leghorn chickens (a total of 40 genomes). A total of 63 variant sites were identified. The rate of insertion/deletion was 16%. Three haplotypes were specified and all the variant sites were at the maximal linkage disequilibrium, i.e. one of the four possible combinations of the alleles at two loci was not found. Two RFLP markers on a bigger sample size confirmed the maximal linkage disequilibrium. The two RFLP gave rise to three alleles (or haplotype classes), which in turn defined six genotypes. A comparison of genotypes revealed significant differences in the body weight, the age of first egg, egg-specific gravity and residual feed consumption. These traits involve many different metabolic pathways, reflecting the major role of ODC in DNA synthesis and cell proliferation.

5.2 INTRODUCTION

Mapping of human disease genes have been shifted from tracing co-inheritance in pedigrees mapping to large scale association studies in population groups (Cardon & Bell, 2001). Such association studies are more powerful for traits which are not inherited in a monogenic fashion (Badano & Katsanis, 2002). In addition, models for statistical analysis require fewer assumptions and may be biologically more realistic (Kuhnlein *et al.*, 2003; Risch, 2000). Efforts are now underway to create a data bank of all human DNA nucleotide polymorphisms (The International SNP Map Working Group, 2001).

DNA polymorphisms are frequent and it has been estimated that in humans two unrelated genomes will differ 1 nucleotide every 1331 bp (Kruglyak & Nickerson, 2001). Hence, typing of every SNP in an association study would be an enormous task and the requirement of multiple tests will reduce the power of the statistical analysis. However, DNA polymorphisms do not segregate independently, but many are physically linked. Hence, a few markers in such linkage groups (i.e. haplotypes) may suffice to characterize the genetic variation of an entire chromosomal region. Even more important, haplotypes provide information about the evolutionary history, including selection pressures on particular chromosomal regions (Gabriel et al, 2002; Tishkoff et al 2001). Haplotype information may be particularly useful for the identification of trait genes in life stock, where the intensive selection is expected to have left its imprint on haplotype structures.

In this communication we characterized the haplotypes of ornithine decarboxylase (ODC) in White Leghorn chickens and analyzed the association of these haplotype with production traits. ODC is the key regulatory enzyme of polyamine synthesis, cations that are important regulators of cell division and differentiation and apoptosis (Thomas &

Thomas, 2001). At the phenotypic level, ODC is elevated in certain types of cancer and is a target for chemotherapy (Russel, 1985, Fisher et al. 2003). In chickens, ODC transcription levels have been associated with growth (Johnson et al, 1995) and with the uptake of calcium in the duodenum (Shinki et al., 1991). These reports suggest that ODC may segregate for natural genetic variants that affect production traits in chickens.

5.3 MATERIALS AND METHODS

5.3.1 Experimental chickens and traits

Chickens were from strain 7 that had been established in Ottawa in 1958 from 4 North American Commercial stocks (Gowe and Fairfull, 1990; Gowe *et al.*, 1993). It consisted of 100 sire families and was maintained by mating one sire with two randomly chosen females. Production data was collected in 1997 from 350 chickens. Each sire family was represented. The maximum number of half-sibs was 6 and of full-sibs was 3. The chickens were weighed at housing (140 days of age, HBWT) and at 265 days of age (MBWT). The onset of laying (AFE) was recorded and rate of egg-laying (HDR) measured between AFE and day 275 (period 1); 276d-356d (period 2) and day 357d-457d (period 3). Egg weight (EWT) and specific gravity (SPG) were measured on eggs laid during 5 consecutive days over three time periods starting from day 240 (period 1), day 350 (period 2) and day 450 (period 3) and averaged for each time period. Assuming that the density of the egg yolk and egg white is one, the increment of SPG over one is approximately equal to the eggshell weight divided by the egg volume and hence a measure of egg-shell strength. Other traits measured were feed consumption, egg mass produced (EGM) and residual feed consumption (RFC) between 247 days and 268 days of age. RFC are the residuals when feed consumption is regressed on the body weight and egg mass produced (Luiting & Urff, 1991).

5.3.2 Identification of DNA variants

DNA was isolated from the blood of 20 male chickens each belonging to a different sire family following standard procedures (Sambrook *et al.*, 1989). Five primer

pairs were designed based on the genomic sequence established by Zhang (1994). They covered most of the region from position 99 to 5074 in five sections. The primer sequences were as follows. Section 1(pos. 99 - 1012): GCT TAC TTT AGC AGT TCT GCG A (forward) and CAG ATA ATT CAC TCT GTG GG AC (reverse). Section 2 (pos. 1063 - 1899) GCT GAC TTT TCC TAG GGT TG (forward) and CAT CTG GAA GAA GCA GAG AC (reverse). Section 3 (pos. 1983 – 2993): GCA CAT GTT AAA CCA GTT CTG CA (forward) and CTC AAT TTA CAC CCA CGA ACA G (reverse). Section 4: (pos. 2982-3992): GTG TAA ATT GAG CTC CGG AC (forward) and TCA AGT GCG GTT AAC AGC AG (reverse). Section 5 (pos. 4023 – 5074): GGA CAG GGA GGA TGA ATT TC (forward) and CAG CCA TTC TAA CAC CAC TC (reverse). The numbering of positions is based on the genomic sequence of Zhang (1994). This sequence starts 327 bp upstream of exon 8 and ends 1159 bp downstream of exon 12. The amplification protocol was 5 min presoaking at 95 C, followed by 35 cycles of 60 sec at 94 C, 80 sec at 55 C and 80 sec at 72 C. The PCR product was purified using a PCR purification kit and sequenced from each end and from internal sequencing primers by the dideoxy method (manufacturer protocol) using an ABI 310 capillary sequencer.

In good quality runs there were no difficulties to identify heterozygote variations involving single base changes. Deletions and insertions lead to the superimposition of two out of register sequences downstream of the deletion/insertion. In most cases this could be resolved by sequencing from the reverse primer or from internal primers. When scoring DNA variants we were conservative and variations that only occurred as single heterozygotes were not scored.

5.3.3 RFLP analysis

The two RFLP analyzed in this study were a *Hind*III RFLP located between exon 11 and 12 at postion 3210 and an *Msp*I RFLP located in the intron preceding exon 8 (about 100 bp upstream from the start of the genomic sequence of Zhang (1994). These RFLP had been initially characterised by Southern blotting (unpublished results). The primers designed subsequently were GCT ATT GTG CTG TCT TGT AAG (forward) and ATG GTA TGC CTC AGC TCA AAT (reverse) for the *Hind*III RFLP and GGC GAT TTC CGA TGC CCG CTG TG (forward) and TCA GAG CCA GGG AAG CCA CCA CC (reverse) for the *Msp*I RFLP. The PCR conditions were the same as for the as described above, except that the annealing temperature was 60 C.

5.3.4 Statistical analysis

The NCSS software used for statistical analysis and for generating the percentile distributions (Hintze, 1997). Association between traits and genotypes were tested using the Kruskal-Wallis analysis of variance by rank and contrasts by the Kruskal-Wallis multiple comparison by the Kruskal-Wallis Z test. Additivity and dominance were analyzed using parametric ANOVA and multi-comparison procedures.

DNAMAN version 5.2.9 (Lynnon Corp.) was used to align sequences and establish gene trees.

5.4 RESULTS AND DISCUSSION

5.4.1. Characterization of ODC haplotypes and analysis of the mutation spectrum

About 4800 bp of the 3'-end of the ODC gene were sequenced in 20 chickens. It had been previously characterized by Zhang (1994) and comprises exon 8 to 12. Among the 20 chickens three different homozygous sequences were observed, thus defining three haplotypes. The remaining sequences were heterozygous but could be explained as combinations of the three haplotypes defined by the homozygous sequences. Genotyping the 20 samples for two RFLP, one located up-stream of exon 8 and the other between exon 11 and 12, indicated that the three haplotypes coincided with three combinations of the four alleles defined by the two RFLP (Table 5.1). The fourth possible combination of RFLP alleles was not observed in this sample nor in 350 individuals analyzed from this strain (Table 5.1).

The sequences of the three haplotypes as well as the sequence of a meat-type bird used as a standard are shown in the appendix and sequence variations are summerized in Table 5.2. Among the 629 bp of exon sequenced only one sequence difference, a C/T transition in exon 11, was observed. The mutation was in the third codon for asparagine and hence did not affect the amino acid sequence of ODC.

In contrast to the exons, there was a considerable variation among intron sequences. Among the 4025 bp sequenced 62 variations were observed. Among these 16% were deletion/insertions, 59% transitions and 25% transversions (Table 5.2). Transitions were more frequent than transversions, as known for other species. However, the high frequency of deletion/insersion was surprising. In comparison, analysis of 3800 base pairs of PEPCK-C in 32 White Leghorn chickens revealed 19 single base changes

and not a single deletion/insertion event (Parsanejad *et al.* 2002). Another striking difference was that despite the large number of *polymorphisms*, only three haplotypes were observed, whereas the 19 SNPs in PEPCK-C gave rise to 6 haplotypes. The numerous sequence differences indicate that three ODC haplotypes have segregated early in the evolution of the chicken. Further, there must have been evolutionary pressures responsible for the maintenance of these early lineages of the ODC gene.

An unrooted tree of the ODC haplotypes in our White Leghorn Strain and the meat-type sequence (haplotype D) used as a reference is shown in Fig. 5.1. There are two internal nodes, one node giving rise to the meat-type and the A haplotype lineages. The other node gives rise to the B and C haplotype lineage. The distance between the two internal nodes is short, indicating the four haplotypes have separated within a short evolutionary time span. Of the three White Leghorn haplotypes, haplotype A is most distant from the meat chicken haplotype, followed by the B and C haplotypes. Further, the branch lengths from the internal node is largest for the D and A haplotypes, respectively. It may indicate selection pressure on these alleles during evolution of the meat-type and egg layer strains, respectively.

5.4.2. Trait association of ODC genotypes

We typed 350 female chickens of strain 7 for the two RFLP, which in the subset of 20 chickens had been diagnostic for the three haplotypes A, B and C. An analysis of the trait associations of the six ODC genotypes is shown in Table 5.3. Significant associations were observed for the body weight at housing, age at first egg, egg specific

gravity and residual feed consumption. The wide variety of traits affected reflects the importance of ODC in many cellular processes.

Contrasts between genotypes indicate that the three ODC alleles are not equivalent. For instance, genotypes AC and AB have significantly different AFE values, indicating that alleles C and B are differ from each other. On the other hand, AA and AB are associated with different SPG1 values, indicating that A and B are also different from each other (Table 5.3).

A comparison of the additive and dominance values for pairs of genotypes further illustrates the different properties of the three alleles (Table 5.4). The homozygote CC has a later AFE and a higher RFC value than the homozygotes AA and BB. This indicates that in this genotype the onset of ovulation is delayed, but it has a better feed utilization. Heterozygotes did not deviate significantly from the average of the homozygotes. Hence, the genotypic effect was additive. In contrast, the SPG of the heterozygotes AB in period 1 was significantly lower than of the homozygotes AA and BB, indicating overdominance. Specific gravity is an approximate measure of the ratio between the eggshell weight and egg volume and hence is strongly linked to Ca metabolism, as well as the mechanism of ovulation. Dominance may reflect that different alleles are optimal for different parts of this pathway.

Both nonparametric statistics and ANOVA analyse differences in the central tendency of distributions. However, traits are determined by a complex network of genes. As a consequence, genetic variations may affect not only the means or medians, but also the shape of trait distributions. Such effects can be graphically illustrated by percentile plots. A percentile plot of SPG (per. 1) for the three genotypes determined by the alleles

A and B of ODC is shown in Fig. 5.2. It indicates that below the median, the distributions for the AA and BB genotypes are identical, whereas the genotype AB is overdominant. However, above the median genotype the percentiles for the BB genotypes are close to the percentiles of the AB genotype AB and above 80^{th} percentile, the percentile values for all three genotypes are similar. Fig. 5.2 also shows the distribution for late SPG, where all three distributions are identical, indicating the age dependency of the ODC genotype on SPG.

A similar effect is observed for AFE (Fig. 5.3). The genotype CC is dominant for a delayed on-set of egg-laying, but the effect is more pronounced below the median than above. Hence it could be that ODC only affects AFE in chickens which start laying eggs early, but has no effect in chickens that start laying eggs late due to variations in other genes.

5.5. CONCLUSION

Polyamines are ubiquitous to every cell and play pivotal role in virtually most signal transduction pathways. It is therefore surprising that chickens that have been under intensive selection since domestication several thousand years ago still segregate for variants in ODC that affect major production traits. This must reflect that epistasis within and between genes provides sufficient buffering to maintain several alleles in the breeding population.

The traits analysed here were complex traits that are determined by the expression patterns of many different genes. Such analyses shed little light on the direct biological mechanism by which allelic variants affect one of the many pathways that lead to phenotypic change. Nevertheless, Shinki *et al.* (1991) have shown that inhibitors of ODC decrease the duodenal content of putrescine and calcium transport activity in new born and five-week chicks and shorten of the length of the duodenal villi. These effects can be reversed by polyamine supplements. Hence, it appears that the lining of the intestine with its rapidly dividing cells is particularly sensitive to fluctuations in ODC activity. More recently, an ODC polymorphism associated with the incidence of colon adenomas has been identified in humans (Martinez *et al.*, 2003). ODC variants therefore may affect the efficiency of nutrient absorption (including Ca), and hence influence traits such as body weight, egg specific gravity and residual feed consumption. AFE may be affected since dietary restrictions are known to delay onset of sexual maturity (Gowe *et al.*, 1960).

Markers that affect production traits can be used to conduct DNA based selection. It should be noted that such selection is not equivalent to phenotypic selection for single traits. Phenotypic selection will affect many different genes, while genotypic selection

will affect several different phenotypes. Variants of ODC were found to affect AFE, SPG, BWT and RFC. On the other hand, a previous analysis of PEPCK-C in the same strain revealed associations with feed efficiency and RFC.

Despite the large number of DNA polymorphisms in the ODC gene, only three haplotypes were observed among 40 genome equivalents. The small number of haplotypes differing by a large number of polymorphisms indicates the presence of a past genetic bottleneck leading to a survival of only a few alleles. PEPCK-C (Parsanejad *et al.*, 2002) and growth hormone (unpublished results) segregate for many more haplotypes in the same strain. Hence the bottleneck does not reflect the founding history of the particular strain. Rather, it may be due to the intensive selection for egg production during the domestication of the chicken. Such selection is expected to leave a signature in the abundance and diversity of haplotypes within and in flanking regions of genes, as has been observed in humans (Tishkoff *et al.*, 2001).

5.6. ACKNOWLEDGMENTS

Research support from the National Science and Engineering Council of Canada, Agriculture and Agri-Food Canada and Shaver Poultry Breeding Farms Ltd. is gratefully acknowledged.

Table 5.1. Distribution of haplotypes in strain 7

Number of chickens	Haplotype	*Hind*III RFLP	*Msp*I RFLP
6	A/A	+/+	+/+
2	B/B	–/–	+/+
1	C/C	–/–	–/–
5	A/B	+/–	+/+
4	A/C	+/–	+/–
2	B/C	–/–	+/–

Table 5.2. Sequence variations in introns

Intron			Transition		Transversion				Del/Ins	Total
Number	Length	Sequenced	C/T	A/G	A/T	A/C	G/T	G/C		
7	-	228	0	1	1	0	1	0	1	4
8	1233	1184	1	4	2	0	1	1	2	11
9	91	91	0	0	0	0	0	0	0	0
10	254	254	1	3	0	1	1	1	3	10
11	1887	1856	17	9	5	0	1	1	4	37
All	-	3613	19	17	8	1	4	3	10	62

Table 5.3. Trait medians for different ODC genotypes.

Trait[1]	ODC Genotype[2]						P-value[3]
	AA (90-78)	AB (87-70)	AC (71-66)	BB (30-25)	BC (38-32)	CC (20-18)	
AFE (d)	167	164	167	168	166	174	**0.011**[4]
BWT (kg)							
130 d	1.28	1.30	1.27	1.29	1.34	1.28	**0.036**
265 d	1.78	1.76	1.74	1.76	1.75	1.64	0.67
Rate of egglaying (%)							
Period 1	85.1	84.8	84.3	82.4	83.9	86.2	0.41
Period 2	71.3	71.3	71.3	73.8	70.0	68.8	0.99
Period 3	57.5	55.0	56.3	61.2	61.9	56.3	0.97
EWT (g)							
Period 1	52.5	52.3	52.3	54.0	51.8	52.7	0.20
Period 2	58.0	58.0	57.7	61.5	57.1	57.5	0.062
Period 3	60.0	60.0	61.1	64.0	60.7	59.7	0.18
SPG[4]							
Period 1	86.9	83.8	84.6	86.0	86.0	86.0	**0.0057**
Period 2	82.0	82.0	81.6	80.7	82.5	82.0	0.34
Period 3	78.0	78.5	78.0	78.0	80.8	80.0	0.50
Feed Cons. (g/day/hen)	117	115	114	120	113	110	0.084
Egg mass (g/day)	46.2	47.0	47.1	47.0	45.9	47.1	0.77
RFC (g)	0.84	0.63	-2.35	5.17	-1.79	-6.68	**0.0060**

[1] The rate of egg laying in three different time period was also analyzed. Its associations with ODC genotypes was not significant. Two outliers with AFE > 250 d and one outlier with a rate of egg laying in period 1 < 50% were removed from the data set.

[2] The number of observations in each group is shown below the ODC genotype. It varied for different traits.

[3] Significance of associations was analyzed by the Kruiskal-Walis analysis of variance by rank, a test we preferred since it does not require normality. However, ANOVA gave the same result.

[4] The following contrasts were significant. AFE: CC>AA, AB, AC, BC and AC>AB. BWT Period 1: AC < AB, BC and AA < BC. SPG1: AA>AB, AC. EWT2: BB> AA, AB, AC, CC. RFC: BC<AA, AB, BB and CC< AA.

Table 5.4: Additivity and dominance of ODC alleles[1]

Trait	Allele Contrast	Additivity Value[2]	Dominance Value[3]
AFE (d)	A, B	1.9	- 1.4
	A, C	8.0***	- 1.9
	B, C	6.1*	- 3.9
SPG Per. 1	A, B	− 1.62	− 1.83**
	A, C	− 1.34	− 0.98
	B, C	0.28	0.86
RFC (g)	A, B	0.7	− 1.2
	A, C	− 6.6**	1.7
	B, C	− 7.3**	− 1.8

[1] Analysis was by ANOVA and homozygotes and heterozygotes were compared using planned comparison procedures. Residuals were normally distributed for all comparisons and traits except for body weight which required log transformation for normalcy. For body weight, additive and dominance values were not significantly different from 0 (data not shown).

[2] Difference between the means of the homozygotes

[3] Difference between the mean of the heterozygotes and the average of the homozygotes

*,**,*** P-value less than 0.05,0.01 and 0.001 respectively.

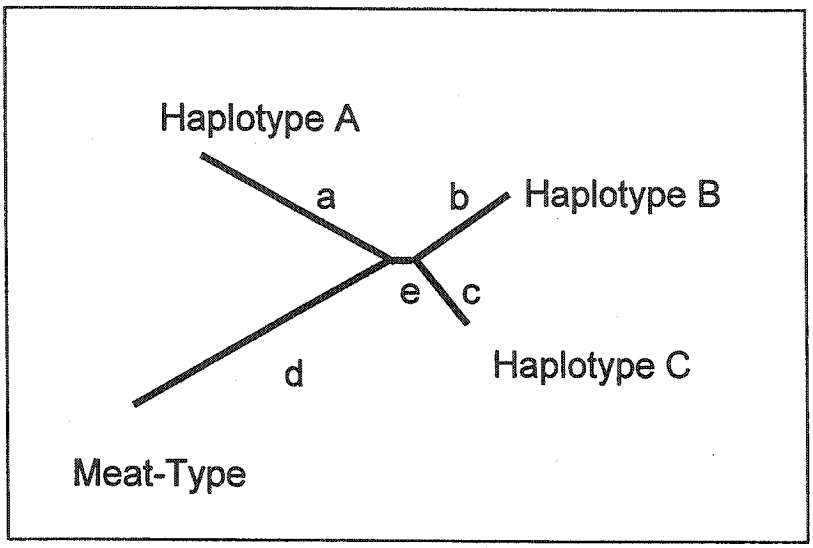


Fig. 5.1: Phylogeny tree of the ODC haplotypes. The gene tree was established using the algorithm by Kimura. However, identical results were obtained with other methods. The length of the branches were a=0.00753, b=0.00062, c=0.00413 and d=0.01015.

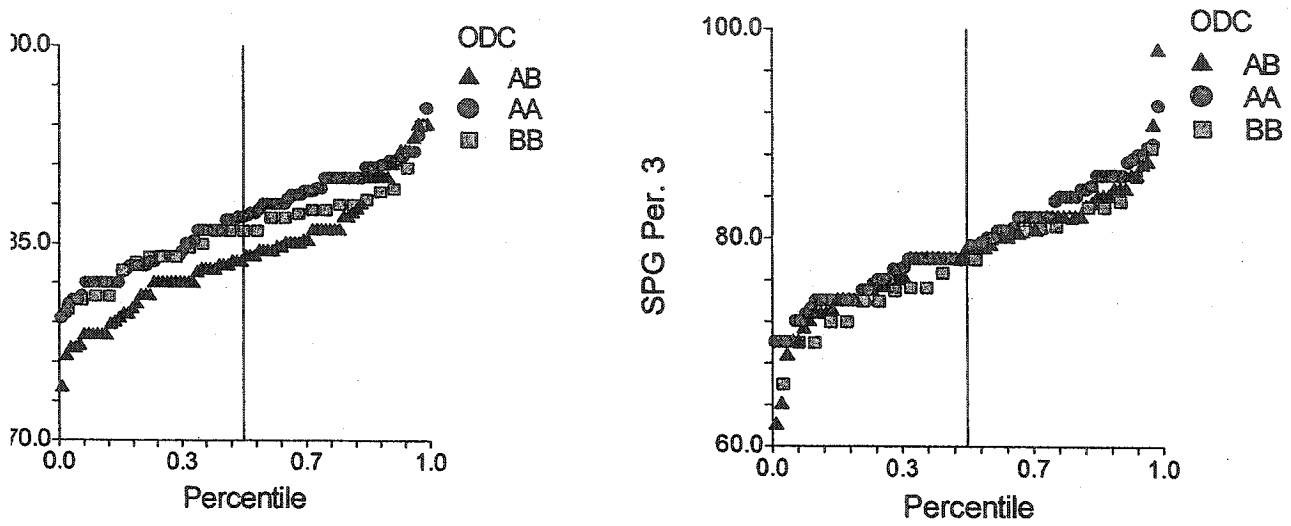


Fig.5.2. Distribution of the percentiles of SPG in periods 1 and 3 for the genotypes defined by the A and B haplotypes. In period 1, the percentiles for the homozygotes AA and BB are nearly identical, but differ above the median. The percentiles for the heterozygote AB are lower than for both homozygotes, except above the 80th percentile. In period 3 the distribution of the percentiles for the three genotypes are nearly identical.

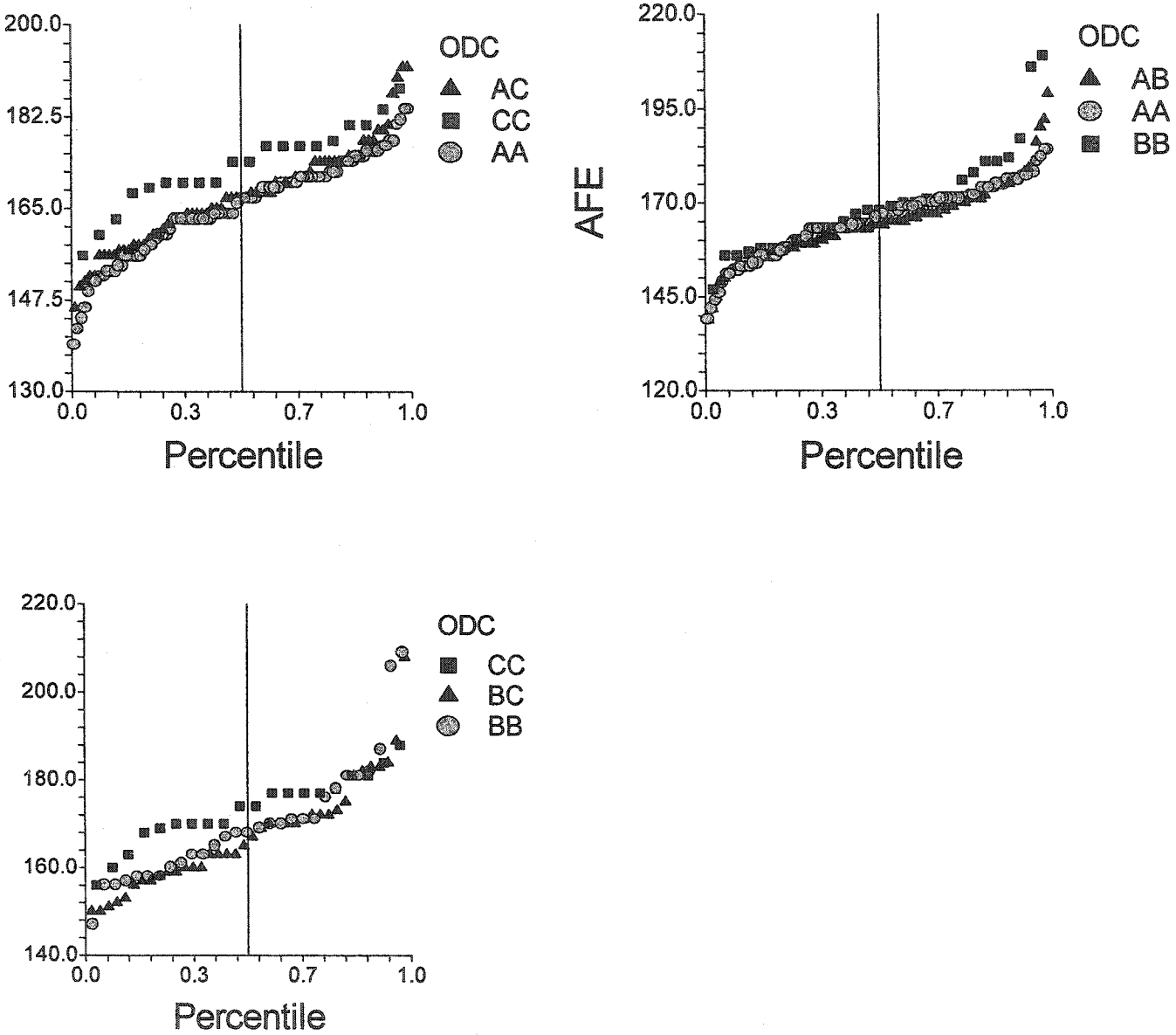


Fig. 5.3. Distribution of the percentiles of AFE

CHAPTER 6

GENERAL CONCLUSION

The aim of this thesis was to identify genetic markers that are associated with major metabolic traits. Rather than screening randomly distributed genetic markers, we analyzed markers in candidate genes. Although this approach does not lead to the discovery of new genes, it addresses two important questions. First, do major regulatory genes segregate for variants that affect gene activity? Second, what is the spectrum of phenotypic traits affected by such variants? Effects of variants that are present in natural breeding populations may differ from those that are introduced artificially and may be relevant in our understanding of gene function and evolution.

The first gene we chose was the PEPCK gene, a key regulatory gene in the pathway leading to de novo synthesis of glycerol and glucose. Glucose and glycerol (via its involvement in lipid biosynthesis) are important molecules in the exchange of metabolic energy between organs. The second gene, ODC, is a key regulatory gene in the synthesis of spermidine, spermine and putrescine. These polyamines modulate DNA-protein and protein–protein interactions and hence interface with most of the intracellular signal pathways. The traits we analyzed in chickens were complex traits that reflect the major input and output of metabolic energy (body weight, feed intake, metabolic rate and calcium production). Any variations that affect the activity of PEPCK or ODC are expected to lead to a phenotypic change in at least one of these traits.

When searching for DNA variants, sequence analysis is the most informative method. It provides complete knowledge of the type, position and context of every

variation, regardless of whether it is a single nucleotide substitution or insertion/deletion variation. Sequence analysis also provides information about the phase between the alleles of marker loci. This information can be used to delineate the haplotypes that segregate in a particular strain and study the evolutionary relationship between these haplotypes. Once the major haplotypes have been identified, markers that are diagnostic for the haplotypes can be identified and used to rapidly genotype a large number of individuals. As an example, the 19 polymorphisms in PEPCK-C about 4 kb led to only 6 haplotypes. These six haplotypes can be identified by five diagnostic SNP. The efficiency of using this method for ODC was even higher. We found 63 polymorphisms in around 5 kb of the genomic DNA giving rise to only 3 haplotypes that could be identified by 2 SNP.

A sequence run is about 500 bp in length. It is therefore quite labour intensive to sequence an entire gene and a decision has to be made about the region of the gene to be sequenced. Conserved regions (exons and regulatory regions) may harbour variations with phenotypic consequences. However such sequences and their location are unknown and may be scattered within the gene. It may therefore be better to first focus on several short intronic regions and choose the most variable one for further sequence analysis and the construction of a gene tree. Although we have not done it here, it may be useful to sequence regions that precede or follow the gene. The phase between the markers in these flanking regions is indicative for intragenic recombination or gene conversion. Once the haplotype structures have been ascertained, different haplotypes may be sequenced in their entirety when searching for potential functional mutations.

The frequency of DNA variants in the ODC gene was about three fold higher than the PEPCK-C gene. In addition, 16% of all the variants in ODC were insertion/deletion

insertion/deletion events while all of the SNPs in PEPCK-C were base substitutions. There are several possibilities for such differences in the mutation distribution. It may be that the three ODC alleles are very ancient, but have gone through a genetic bottleneck at a relatively recent time in evolution. PEPCK-C on the other hand may have descended from a single more recent allele, but may subsequently not have been under stringent selection. Alternatively, the chromatin structures may be different for the two genes, resulting in differences in the rate and type of mutations.

The large difference in the mutation frequency might reflect also that different regions of the two genes were sequenced. The sequenced regions in the ODC gene were internal introns, whereas, most of the non-coding sequence of the PEPCK-C sequence covered the 5'-flanking promoter region. The 5'flanking region may be more conserved than internal introns. However, even among the internal introns of ODC, there were significantly different mutation rates. Further, conservation of sequences could hardly explain the abundance of deletion/insertions in ODC.

The traits we have analyzed are major energy input and output traits that are determined by many different genes. Genetic variants that affect regulatory genes of basic metabolic pathways are expected to affect all of these traits. However, PEPCK-C only affects feed efficiency and residual feed consumption, while ODC affects body weight, onset of sexual maturity and egg shell calcium. Hence the two spectra of traits have only the residual feed consumption (a measure of metabolic efficiency) in common. It will be a great challenge to understand how this communality relates to changes by network of genes interaction. Such insights may be gained by analyzing the association of genotypes with the expression pattern of all cellular genes using microarray technology. The

dissection of the mechanisms that link variants in a single gene with a complex trait requires also the establishment of data sets that include more less complex subtraits.

For both, PEPCK-C and ODC we have emphasized that establishing the haplotype structure of a gene greatly reduces the amount of genotyping necessary to characterize genetic variants. A further premise was that in general a gene segregates for a single QTL. A gene tree established from the haplotype analysis could then be used to map the QTL to a specific branch of the tree. However this was not the case. Both, ODC and PEPCK-C segregated for more than one QTL and each one of these QTL was associated with different phenotypic properties. Retroactively, this may not be surprising. If an organism tolerates variation in the activity of gene at all, it may also tolerate other types of biologically consequential mutations.

In view of the intensive selection exercised in chickens, it is surprising that genes segregating for major QTL are still present. Such genes may have been maintained due to overdominance (i.e. the hybrid may be superior to the two homozygotes). Overdominance for eggshell calcium was indeed observed for one of the ODC alleles. Alternatively, alleles may be maintained due to interaction with other genes. PEPCK-C was shown to interact with PEPCK-M. In particular, the effect of the genotype of one gene was entirely dependant on the genotype of the second gene.

Whether a genotype survives in the population depends on its effect on fitness. For the chickens, the most important parameters are the number of eggs laid, fertility, hatchability, mortality and selection by the breeder. Differential fitness is expected to lead to Hardy-Weinberg disequilibrium, even if the population is randomly mated. This is indeed observed for PEPCK-C. Although we have shown that PEPCK-C might be rate limiting in chickens with high, but not with low body weight, it is unlikely that the

reduction in housing body weight is directly associated with reduced reproductive fitness. Rather, PEPCK-C in affiliation with other genes may affect fitness. Deviation from Hardy-Weinberg equilibrium was only observed for PEPCK-C but not a series of other genes that affect production traits.

Disequilibrium analysis can be extended to intragenic disequilibria. In particular, one can analyze whether the combination of alleles at two different loci deviate from random segregation. Such analyses may involve two, three or four alleles, leading to the definition of digenic, trigenic and quadrigenic disequilibria. PEPCK-C and PEPCK-M genotypes were at significant trigenic disequilibrium. A comparison of the genotypic combinations at an excess to those at a deficit revealed a significant association with egg shell calcium, a trait that may possibly related to fitness.

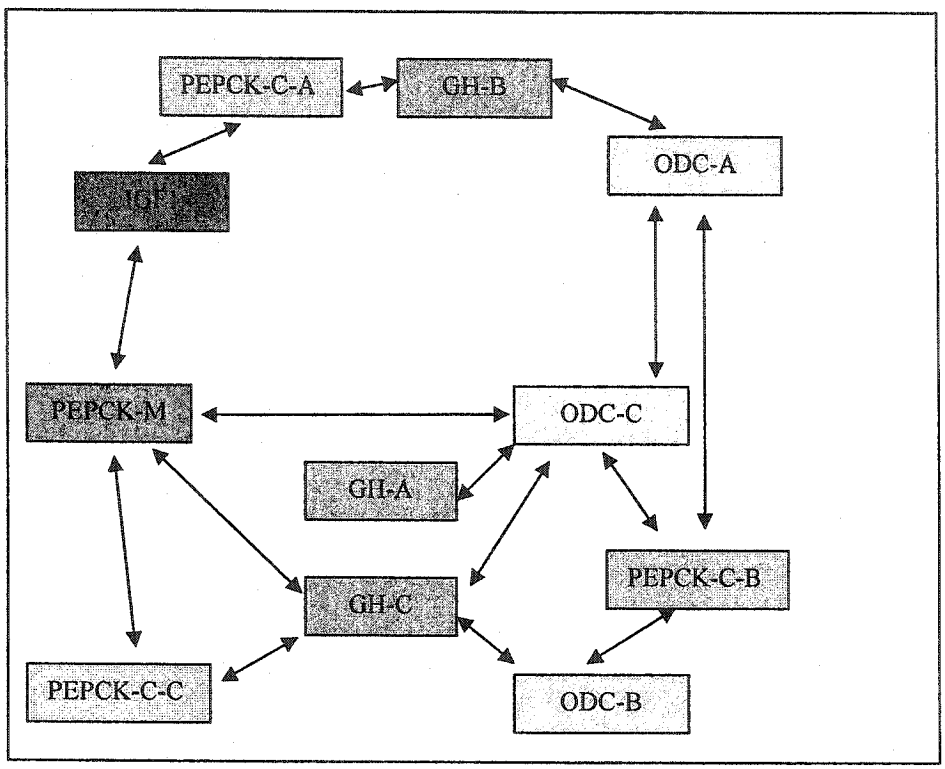
The disequilibrium between PEPCK-C and PEPCK-M is not an exception but it can be observed with many other genes. Kuhnlein *et al.* (2003) examined the genotypic disequilibrium between pairs of 5 genes segregating for 13 alleles. (Figure 6.1). They showed that all 5 genes segregated for alleles giving rise to digenic or higher order linkage intergenic disequilibrium. Among a total of 67 pairs, 14 pairs gave rise to genotypes at disequilibrium. The actual number may be higher, since the frequency of some genotypes may be too low to reveal significant deviation from equilibrium.

In addition to ODC, PEPCK-C and PEPCK-M have been analyzed the GH-receptor, GH, IGF1 and the mitochondrial genome in the same strain. All of these genes segregate for alleles that affect major metabolic traits in chickens. Although we have chosen these genes based on their metabolic role, it indicates that QTL are frequent. Similar observations were made by Dr. Lamont (personal communication to Urs Kuhnlein). On the other hand, linkage mapping in crosses of inbred lines yielded fewer

QTL. This is partially due to gene interaction that renders the phenotypic effect of a variant dependent on the presence of variations in other genes.

The interactive nature of QTL indicates that phenotypes are determined by nonlinear pathways composed of numerous branches and loops. To understand such pathways we require detailed analyses of the interrelationship between variants and changes in the expression of the entire organism. Such analyses will keep scientists busy for many decades. Apart from providing basic insights into the relation between genes and phenotypes, QTL analysis provides a tool for DNA based selection. At present it is difficult to predict the outcome of such selection. The consequence of such selection on other traits and the long-term effect on the frequency of variants in other genes have to be determined empirically.

The chicken genome is currently being sequenced and genetic study in chickens where large databases are available will become really exciting.



Adapted from Kuhnlein *et al.*, 2003

FIGURE 6.1. Genotypic disequilibria between alleles at 6 loci in strain 7. GH: alleles are combinations of a *Sac*I RFLP and a *Msp*I RFLP (Kuhnlein *et al.*, 1997). ODC: alleles are combinations of a *Hind*III RFLP and a *Msp*I RFLP (Parsanejad *et al.*, 2003b). PEPCK-C: alleles are combinations of an *Aci*I RFLP and a *Bst*EII RFLP (Parsanejad *et al.*, 2002, 2003a). The PEPCK-M: alleles are defined by an *Acc*I RFLP (Parsanejad *et al.*, 2003a). IGF1: alleles are defined by a *Pst*I RFLP (Nagajara *et al.*, 2000). Arrows connect alleles with significant digenic, trigenic or quadrigenic disequilibrium coefficients (Weir, 1990; Lewis and Zaykin, 2001).

REFERENCES

Abecasis, G.R., L.R. Cardon and W.O. Cookson. 2000. A general test of association for quantitative traits in nuclear families. Am. J. Hum. Genet. 66(1):279-92.

Aggrey, S.E., M. Lessard, D. Hutchings, S. Josheph, X.P. Feng, D. Zadworny, and U. Kehnlein. 1996. Association of genetic markers with immune traits, The 5th international sympossium on Marek's disease. Page 80-85.

Alhonen, L., J.J. Parkkinen, T. Keinanen, R. Sinervirta, K.H. Herzig and J. Janne. 2000. Activation of polyamine catabolism in transgenic rats induces acute pancreatitis. Proc. Natl. Acad. Sci. USA 97:8290–8295.

Alm, K., P.S. Berntsson, D.L. Kramer, C.W. Porter and S.M. Oredsson. 2000. Treatment of cells with the polyamine analog N1, N11-diethylnorspermine retards S phase progression within one cell cycle. Eur. J. Biochem. 267: 4157–4164.

Altshuler, D., V.J. Pollara, C.R. Cowles, W.J. Van Etten, J. Baldwin, L. Linton, and E.S. Lander. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 407(6803):513-6.

Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet. 2(2):130-8.

Badano, J.L. and N. Katsanis. 2002. Beyond Mendel: an evolving view of human genetic disease transmission. Nat. Rev. Genet. 3(10):779-89.

Ballard, F.J., R.W. Hanson and G.A. Leveille. 1967. Phosphoenolpyruvate carboxykinase and the synthesis of glyceride-glycerol from pyruvate in adipose tissue. J Biol Chem. 242(11):2746-50.

Beale, E.G., N.B. Chrapkiewicz, H.A. Scoble, R.J. Metz, D.P. Quick, R.L. Noble, J.E. Donelson, K. Biemann and D.K. Granner. 1985. Rat Hepatic Cytosolic Phosphoenolpyruvate Carboxykinase (GTP). Journal of Biology Chemistry 260:10748-10760.

Bisbis, S., M. Derouet and J. Simon. 1994. Characterization of insulin receptors in chicken kidneys: effect of nutritional status. Gen Comp Endocrinol. 96(1):37-49.

Cardon L.R. and J.L. Bell, 2001. Association study designs for complex diseases. Nat Rev Genet. 2(2):91-9.

Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie, N. Patil, N. Shaw, C.R. Lane, E.P. Lim, N. Kalyanaraman, J. Nemesh, L. Ziaugra, L. Friedland, A. Rolfe, J. Warrington, R. Lipshutz, G.Q. Daley, and E.S. Lander. 1999. Characterization of single-nucleotide polymorphism in coding regions of human genes. Nat. Genet. 22:231-238.

Chicurel, M. 2001. Faster, better, cheaper genotyping. Nature. 412(6847):580-2.

Collins, S.C., M.S. Guyer, and A. Chakravarti. 1997. Variations on a theme: Cataloging human DNA sequence variation. Sci. 278:1580-1581.

Cook, J.S., S.L. Weldon, J.P. Garcia-Ruiz, Y. Hod and R.W. Hanson. 1986. Nucleotide Sequence of the mRNA Encoding the Cytosolic Form of Phosphoenolpyruvate Carboxykinase (GTP) from the Chicken. Proc. Natl. Axad. Sci. USA. 83:7583-87.

da Rocha, A.B., D.R. Mans, A. Regner, G. and Schwartsmann. 2002. Targeting protein kinase C: new therapeutic opportunities against high-grade malignant gliomas? Oncologist. 7(1):17-33.

Daly, M.J., J.D. Rioux, S.F. Schaffner, T.J. Hudson and E.S. Lander. 2001 High-resolution haplotype structure in the human genome. Nat Genet. 29(2):229-32.

Dawson, E., G.R. Abecasis, S. Bumpstead, Y. Chen, S. Hunt, D.M. Beare, J. Pabial, T. Dibling, E. Tinsley, S. Kirby, D. Carter, M. Papaspyridonos, S. Livingstone, R. Ganske, E. Lohmussaar, J. Zernant, N. Tonisson, M. Remm, R. Magi, T. Puurand, J. Vilo, A. Kurg, K. Rice, P. Deloukas, R. Mott, A. Metspalu, D.R. Bentley, L.R. Cardon and I. Dunham. 2002 A first-generation linkage disequilibrium map of human chromosome 22. Nature. 418(6897):544-8.

Dermitzakis, E.T., A. Reymond, R. Lyle, N. Scamuffa, C. Ucla, S. Deutsch, B.J. Stevenson, V. Flegel, P. Bucher, C.V. Jongeneel, S.E. Antonarakis.2002. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature. 420(6915):578-82.

Devlin, B, K. Roeder, L. and Wasserman. 2001. Genomic control, a new approach to genetic-based association studies. Theor Popul Biol. 60(3):155-66.

Dopazo, J, E. Zanders, I. Dragoni, G. Amphlett and F. Falciani. 2001. Methods and approaches in the analysis of gene expression data. J Immunol Methods. 250(1-2):93-112.

Doris, P.A. 2002. Hypertension genetics, single nucleotide polymorphisms, and the common disease: common variant hypothesis. Hypertension. 39(2 Pt 2):323-31.

Eddy, S.R. 2001. Non-coding RNA genes and the modern RNA world. Nat Rev Genet. 2(12):919-29.

Falaki, M., N. Gengler, M. Sneyers, A. Prandi, S. Massart, A. Formigoni, A. Burny, D. Portetelle and R. Renaville. 1996. Relationships of polymorphisms for growth hormone and growth hormone receptor genes with milk production traits for Italian Holstein-Friesian bulls. J Dairy Sci. 79(8):1446-53.

Fan, J.B., X. Chen, M.K. Halushka, A. Berno, X. Huang, T. Ryder, R.J. Lipshutz, D.J. Lockhart and A. Chakravarti. 2000. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. Genome Res.10(6):853-60.

Feng, X.P., U. Kuhnlein, R.W. Fairfull, S.E. Aggrey, J. Yao J and D. Zadworny. 1998. A genetic marker in the growth hormone receptor gene associated with body weight in chickens.J Hered. 89(4):355-9.

Feng, X.P., U. Kuhnlein, S.E. Aggrey, J.S. Gavora and D. Zadworny. 1997. Trait association of genetic markers in the growth hormone and the growth hormone receptor gene in a White Leghorn strain. Poult Sci. 76(12):1770-5.

Fischer, S.M., C.J. Conti, J. Viner, C.M. Aldaz and R.A. Lubet. 2003 Celecoxib and difluoromethylornithine in combination have strong therapeutic activity against UV-induced skin tumors in mice. Carcinogenesis. 24(5):945-52.

Fredlund, J.O. and S.M. Oredsson.1996. Normal G1/S transition and prolonged S phase within one cell cycle after seeding cells in the presence of an ornithine decarboxylase inhibitor. Cell Prolif. 29(8): 457-66.

Gabriel, S.B., S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly and D. Altshuler. 2002. The structure of haplotype blocks in the human genome. Science. 296(5576):2225-9.

Gowe R.S. and R.W. Fairfull. 1990. Genetic controls in selection. Page 935-954 in: Poultry breeding and genetics. R.D. Crawford, ed. Amsterdam-Oxford-New York-Tokyo: Elsevier.

Gowe, R. S., A. S. Johnson, R. D. Crawford, J. H. Downs, A. T. Hill, W. F. Mountain, J. R., Pelletier, and J. H. Strain. 1960. Restricted versus full feeding during the growing period for egg production stock. Brit. Poult. Sc. 1:37-56

Gowe, R.S., R.W. Fairfull, I. McMillan and G.S. Schmit. 1993. A strategy for maintaining high fertility and hatchability in a multiple-trait egg stock selection program. Poult. Sci. 72:1433-1448.

Hacia, J.G., J.B. Fan, O. Ryder, L. Jin, K. Edgemon, G. Ghandour, R.A. Mayer, B. Sun, L. Hsie, C.M. Robbins, L.C. Brody, D. Wang, E.S. Lander, R. Lipshutz, S.P. Fodor, and F.S. Collins. 1999. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat. Genet. 22:164-167.

Halushka, M.K., J.B. Fan, K. Bentley, L. Hsie, N. Shen, A. Weder, R. Cooper, R. Lipshutz, and A. Chakravarti. 1999. Patterns of Single-Nucleotide Polymorphisms in Candidate Genes for Blood-Pressure Homeostasis. Nat. Genet. 22: 239-247.

Hanson, R.W. and L. Reshef. 1997. Regulation of phosphoenolpyruvate carboxykinase (GTP) gene expression. Annual Review of Biochemistry 66:581 - 611.

Hanson, R.W. and Y.M. Patel. 1994. Phosphoenolpyruvate Carboxykinase (GTP): The Gene and the Enzyme. Advance of Enzymology. 69:203-281.

Hartman, J.L., B. Garvik, and L. Hartwell.2001. Principles for the buffering of genetic variation. Sci. 291(5506):1001-4.

Hernandez, J.L.and B.S. Weir. 1989. A disequilibrium coefficient approach to Hardy-Weinberg testing. Biometrics. 45(1):53-70

Hickok, NJ, J. Wahlfors, A. Crozat, M. Halmekyto, L. Alhonen, J. Janne and O.A. Janne. 1990. Human ornithine decarboxylase-encoding loci: nucleotide sequence of the expressed gene and characterization of a pseudogene. Gene. 93(2):257-63.

Hintze J.L., 1997. Statistical system for windows. Number cruncher statistical systems, Kaysville, Utah.

Hod, Y., H.Y. Warren and R.W. Hanson. 1984. The Gene Encoding the Cytosolic Form of Phosphoenolpyruvate Carboxykinase (GTP) from the Chicken. Journal of Biological Chemistry. 259(24):15609–15614.

Hoet, P. H. M. and B. Nemery. 2000. Polyamines in the lung: polyamine uptake and polyamine-linked pathological toxicological conditions. Am. J. Physiol. Lung Cell Mol. Physiol. 278: L417–L433

Johnson, M.I. 1984. The World and the Sickle Cell Gene: A Study in Health Education. Trado-medic Books.

Johnson, R. and G. Bulfield. 1992. Molecular cloning and sequence analysis of a chicken ornithine decarboxylase cDNA. Anim Genet. 23(5):403-9.

Johnson, R., G. Bulfield, A. Tait, and C. Goddard. 1995. Increased expression and activity of ornithine decarboxylase in chicks genetically selected for rapid growth rate. Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 110(3):531-7.

Kajiwara, K, E.L. Berson and T.P Dryja. 1994. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci.Sci. 264(5165):1604-8.

Kennedy, B.W., M. Quinton and J.A. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. J Anim Sci. 70(7):2000-12.

Konno, R. 2001. Gene expression profiling of human ovarian epithelial tumors by digo nucleotide microarray. Hum Cell. 14(4):261-6.

Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Natu. Gene.22:139-144.

Kruglyak, L., D.A. Nickerson. 2001. Variation is the spice of life. Nat Genet. 27(3):234-6.

Kuhnlein, U., L. Ni, S. Weigend, J.S. Gavora, W. Fairfull, and D. Zadworny. 1997. DNA polymorphisms in the chicken growth hormone gene: response to selection for disease resistance and association with egg production. Anim Genet. 28(2):116-23.

Kuhnlein, U., R. Parsanejad, D. Zadworny, and S. E. Aggrey. 2003. The dynamics of the genotype-phenotype association. Poult Sci. 82(6):876-81.

Kusunoki, S. and I. Yasumasu. 1978. Inhibitory effect of a-hydrazinoornithine on egg cleavage in sea urchin eggs. Dev. Biol. 67: 336–345

Lander E.S. and N.J. Schork. 1994.Genetic dissection of complex traits. Sci. 265(5181):2037-48.

Lander, E.S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121(1):185-99.

Lewis, P. O. and D. Zaykin. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html

Lewontin, R.C. 1988. On measures of gametic disequilibrium.Genet. 120(3):849-52.

Li, S., D. Zadworny, S.E. Aggrey, and U. Kuhnlein. 1998a. Mitochondrial PEPCK: A highly polymorphic gene with alleles co-selected with Marek's disease resistance in chickens. Anim. Genet. 29:395-397.

Li, S., N. Liu, D. Zadworny and U. Kuhnlein. 1998c. Genetic variability in white leghorns revealed by chicken liver expressed sequence tags. Poult Sci. 77(1):134-9.

Li, S., S.E. Aggrey, D. Zadworny, W. Fairfull, and U. Kuhnlein. 1998b. Evidence for a Genetic variation in the Mitochondrial Genome Affecting Traits in White Leghorn Chickens. The Journal of Heredity. 89: 222-226.

Linde, S.M. 1972. Sickle Cell: A complete guide to prevention and treatment. Pavilion Publishing Company.

Luiting, P. and E.M. Urff. 1991. Residual feed consumption in laying hens. 2. Genetic variation and correlations. Poult Sci. 70(8):1663-72

Lweis, P.O. and D. Zaykin. 2001. Genetic data analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html.

Lyamichev, V., A.L. Mast, J.G. Hall, J.R. Prudent, M.W. Kaiser, T. Takova, R.W. Kwiatkowski, T.J. Sander, M. de Arruda, D.A. Arco, B.P. Neri and M.A. Brow.1999. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes.Nat Biotechnol.17(3):292-6.

Martinez, M.E., T.G. O'Brien, K.E. Fultz, N. Babbar, H. Yenushalmi, N. Qu, Y. Guo, D. Boorman, J. Einspahr, D.S. Alberts, and E.W. Gerner. 2003. Pronounced reduction in adenoma recurrence associated with aspirin use and a polymorphism in the ornithine decarboxylase gene. Proc. Natl. Acad. Sci. USA. 100(13):7859-64.

Merila, J. and B.C. Sheldon. 1999. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. Heredity. 83 (Pt 2):103-9.

Morgan T.H., 1919. Drosophila melanogaster. The physical basis of heredity, Philadelphia: J.B. Lippincott Company.

Morgan, D. M. L. 1999. Polyamines: an overview. Mol. Biotechnol. 11: 229–250.

Morgan, R.W., L. Sofer, A.S. Anderson, E.L. Bernberg, J. Cui, and J. Burnside. 2001. Induction of host gene expression following infection of chicken embryo fibroblasts with oncogenic Marek's disease virus. J Virol. 75(1):533-9.

Morisson, M., A. Bordas, J..M. Petit, C. Jayat-Vignoles, R. Julien and F. Minvielle. 1997. Associated effects of divergent selection for residual feed consumption on reproduction, sperm characteristics, and mitochondria of spermatozoa. Poult. Sci. 76(3):425-31.

Mullikin, J.C, S.E. Hunt, C.G. Cole, B.J. Mortimore, C.M. Rice, J. Burton, L.H. Matthews, R. Pavitt, R.W. Plumb, S.K. Sims, R.M. Ainscough, J. Attwood, J.M. Bailey, K. Barlow, R.M. Bruskiewich, P.N. Butcher, N.P.Carter, Y. Chen, C.M. Clee, P.C. Coggill, J. Davies, R.M. Davies, E. Dawson, M.D. Francis, A.A. Joy, R.G. Lamble, C.F. Langford, J. Macarthy, V. Mall, A. Moreland, E.K. Overton-Larty, M.T. Ross, L.C. Smith, C.A. Steward, J.E. Sulston, E.J. Tinsley, K.J. Turney, D.L. Willey, G.D. Wilson, A.A. McMurray, I. Dunham, J. Rogers, D.R. Bentley. 2000. An SNP map of human chromosome 22. Nature. 407(6803):516-20.

Nadeau, J.H., 2001. Modifier genes in mice and humans. Nat Rev Genet. 2(3):165-74.

Nagaraja, S.C., S.E. Aggrey, J. Yao, D. Zadworny, R.W. Fairfull, and U. Kuhnlein. 2000. Trait association of a genetic marker near the IGF-I gene in egg-laying chickens. J Hered. 91(2):150-6.

Ng-Kwai-Hang, K.F., D. Zadworny, J.F. Hayes and U. Kuhnlein. 1991. Identification of kappa-casein genotype in Holstein sires: a comparison between analysis of milk samples from daughters and direct analysis of semen samples from sires by polymerase chain reaction. J Dairy Sci. 74(8):2410-5.

Ohnishi, Y., T. Tanaka, K. Ozaki, R. Yamada, H. Suzuki and Y. Nakamura. 2001. A high-throughput SNP typing system for genome-wide association studies. J. Hum. Genet. 46(8):471-7.

Olswang, Y., H. Cohen, O. Papo, H. Cassuto, C.M. Croniger, P. Hakimi, S.M. Tilghman, R.W. Hanson and L. Reshef. 2002. A mutation in the peroxisome proliferator-activated receptor gamma-binding site in the gene for the cytosolic form of phosphoenolpyruvate carboxykinase reduces adipose tissue size and fat content in mice. Proc Natl Acad Sci U S A.99(2):625-30.

Parsanejad, R., A. Torkamanzehi, D. Zadworny and U. Kuhnlein, 2003. Alleles of Cytosolic PEPCK: Trait Association and Interaction with Mitochondrial PEPCK in a Strain of White Leghorn Chickens. Poult. Sci. 82(11):1708-15.

Parsanejad, R., D. Zadworny, and U. Kuhnlein, 2002. Genetic variability of the cytosolic phosphoenolpyruvate carboxykinase gene in White Leghorn chickens. Poult. Sci. 81(11):1668-70.

Parsanejad, R., D. Zadworny, and U. Kuhnlein,2004, Ornithine decarboxylase: Haplotype Structure and Trait association in white leghorn chickens. Poult. Sci. Accepted.

Pegg, A.E. and P.P. McCann. 1982. Polyamine metabolism and function. Am. J. Physiol. 243: C212–C221.

Rannala, B. 2001. Finding genes influencing susceptibility to complex diseases in the post-genome era. Am J Pharmacogenomics. 1(3):203-21.

Risch, N. and K. Merikangas. 1996. The future of genetic studies of complex human diseases. Sci. 273(5281):1516-7.

Russell, D.H. 1985. Ornithine decarboxylase: a key regulatory enzyme in normal and neoplastic growth. Drug Metab Rev. 16(1-2):1-88.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Page 916-919 in Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Page 916-919 in Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sato, A., H. Takahashi, K. Konishi, T. Suzuki and H. Kochi. 1997. Nucleotide sequence of the promoter region of chicken cytosolic phosphoenolpyruvate carboxykinase gene. J. Biochem. 121:711–716.

Savon, S.P., P. Hakimi and R.W. Hanson. 1993. Expression of the Genes for the Mitochondrial and Cytosolic Form of Phosphoenolpyruvate Carboxykinase (GTP) in Avian Liver during Development. Biological Neonate. 64: 62–68.

Savon, S.P., P. Hakimi, D.R. Crawford, D.J. Klemn, A.L. Gurney and R.W. Hanson. 1997. The Promoter Regulatory Regions of the Genes for the Cytosolic Form of Phosphenolpyruvate Carboxykinase (GTP) from the Chicken and the Rat Have Different Species-Specific Roles IN Gluconeogenesis. Journal of Natr. 127: 276–85.

Seiler, N. 2000. Oxidation of polyamines and brain injury. Neurochem. Res. 25:471–490.

Seiler, N. and C.L. Atanassov. 1994. The natural polyamines and the immune system. Prog. Drug Res. 43:87–141.

She, P., M. Shiota, K.D. Shelton, R. Chalkley, C. Postic and M.A. Magnuson. 2000. Phosphoenolpyruvate carboxykinase is necessary for the integration of hepatic energy metabolism. Mol Cell Biol. 20(17):6508-17.

Shinki, T., H. Tanaka, J. Takito, A. Yamaguchi, Y. Nakamura, S. Yoshiki and T. Suda. 1991. Putrescine is involved in the vitamin D action in chick intestine. Gastroenterology. 100(1):113-22.

Smith, T.K. 1990. Effect of dietary putrescine on whole body growth and polyamine metabolism. Proc. Soc. Exp. Biol. Med. 194(4):332-6.

Tabor, H.K., N.J. Risch and R.M. Myers, 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat. Rev. Genet.;3(5):391-7.

Thomas, T and T.J. Thomas. 2001. Polyamines in cell growth and cell death: molecular mechanisms and therapeutic applications. Cell Mol Life Sci. 58(2):244-58.

Thomas, T and T.J. Thomas.2003. Polyamine metabolism and cancer. J Cell Mol Med. 7(2):113-26.

Thomas, T. J., U.B. Gunnia and T. Thomas. 1992. Reversal of the abnormal development of T cell subpopulations in the thymus of autoimmune MRL-pr:lpr mice by a polyamine biosynthesis inhibitor. Autoimmunity. 13:275–283.

Tishkoff, S.A., R. Varkonyi, N. Cahinhinan, S. Abbes, G. Argyropoulos, G. Destro-Bisol, A. Drousiotou, B. Dangerfield, G. Lefranc, J. Loiselet, A. Piro, M. Stoneking, A. Tagarelli, G. Tagarelli, E.H. Touma, S.M. Williams and A.G.Clark. 2001 Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science. 293(5529):455-62.

Utter, M.F. and K. Kurahashi. 1954. Mechanism of action of Oxaloacetate Carboxylase. Journal of American Soc. 75:785-787.

Vallejo, R.L., L.D. Bacon, H.C. Liu, R.L. Witter, M.A. Groenen, j. Hillel AND H.H. Cheng. 1998. Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease virus induced tumors in F2 intercross chickens.Genet. 148(1):349-60.

Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J.Mural,G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans , R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R Brandon., M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T.

Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N.N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigo, M.J. Campbell, K.V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu. 2001. The sequence of the human genome. Sci. 16;291(5507):1304-51. Erratum in: Sci. 5;292(5523):1838.

Wacholder S, M. Garcia-Closas and N. Rothman. 2002. Study of genes and environmental factors in complex diseases (letter). Lancet. 359:1155

Wacholder S, N. Rothman and N. Caporaso. 2000. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst. 92:1151–1158

Wall, J.D. and J.K.Pritchard. 2003. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 4(8):587-97.

Wanstrat, A. and E. Wakeland. 2001. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol. 2(9):802-9.

Wasserman, R.H. and C.S. Fullmer. 1989. On the molecular mechanism of intestinal calcium transport., Adv Exp Med Biol. 249:45-65.

Weir, B. 1990. Genetic data analysis. Sinauer, Sunderland, MA.

Weiss, K.M. and J.D. Terwilliger. 2000. How many diseases does it take to map a gene with SNPs? Nat Genet. 26(2):151-7. Review.

Weldon, S.L., A. Rando, A. S. Matath, Y. Hod, P.A. Kalovic, S. Savon and J.S. Cook. 1990. Mitochondrial P-enolpyruvate carboxykinase from the chicken. Nucleotide sequence of the cDNA and the deduced amino acid sequence. J. Biol. Chem. 165:7308-7317.

Witmer, P.D., K.F. Doheny, M.K. Adams, C.D. Boehm, J.S. Dizon, J.L. Goldstein, T.M. Templeton, A.M. Wheaton, P.N. Dong, E.W. Pugh, R.L. Nussbaum, K. Hunter, J.A. Kelmenson, L.B. Roweand M.J. Brownstein. 2003. The development of a highly informative mouse Simple Sequence Length Polymorphism (SSLP) marker set and construction of a mouse family tree using parsimony analysis. Genome Res. 13(3):485-91.

Xiong, M., S.W. Guo. 1997. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. Am J Hum Genet. 60(6):1513-31.

Yao, J., D. Zadworny, U. Kuhnlein and J.F. Hayes. 1995. Molecular cloning of a bovine ornithine decarboxylase cDNA and its use in the detection of restriction fragment length polymorphisms in Holsteins. Genome. 38(2):325-31.

Yao, J., S.E. Aggrey, D. Zadworny, J.F. Hayes and U. Kuhnlein. 1996. Sequence variations in the bovine growth hormone gene characterized by single-strand conformation polymorphism (SSCP) analysis and their association with milk production traits in Holsteins.Genetics. 144(4):1809-16.

Yao, J., S.E. Aggrey, D. Zadworny, U. Kuhnlein and J.F. Hayes. 1998. Restriction fragment length polymorphisms at the ornithine decarboxylase locus associated with milk protein yield in Holsteins. J Dairy Res. 65(2):341-5.

Zhang, L. 1994, Molecular cloning and characterization of the chicken ornithine decarboxylase gene. McGill University, M.Sc. thesis.

APPENDIX

Sequence Alignment of ODC gene:

The sequences of the three haplotypes was aligned with the sequence of Zhang (1994). The sequence of Zhang was edited in two instances, where in comparison to the haplotype sequences the position of adjacent nucleotide was switched (i.e. TA to AT). Such errors are expected to occur relatively frequently when sequences are read from autoradiographs. Exons are in bold. When occurring within the reported sequence, the PCR primer sequences are underlined. The sequences of three nucleotide in region 2 and a large T repeat in intron 5 were unclear and are designated by Ns. The sequences of the haplotypes H1, H2 and H3 correspond to the haplotypes A, B and C. The sequence of Zhang (1994) is designated as LODC3' (meat bird).

(A) Sequence Alignment in Region 1

```
ODC1H1    ........................................    0
ODC1H2    ........................................    0
ODC1H3    ........................................    0
LODC3'    gcataatacgactcactatagggatcagttatagaactgt    40

ODC1H1    ........................................    0
ODC1H2    ........................................    0
ODC1H3    ........................................    0
LODC3'    tacagggttaaaagttgaagttcctcatttcaggttagtt    80

ODC1H1    ..................GCTTACTTTAGCAGTTCTGCGA    22
ODC1H2    ..................----------------------    22
ODC1H3    ..................----------------------    22
LODC3'    ttttttgagagcttggtg----------------------    120

ODC1H1    TAAAATCAACAGCTCAAACTGAGATGACTCTTACAAGAAC    62
ODC1H2    ----------------------------------------    62
ODC1H3    ----------------------------------------    62
LODC3'    ----------------------------------------    160

ODC1H1    CCAGTAAATTGGCTCTGTTCTCAAATCA............    90
ODC1H2    ----------------------------g...........    91
ODC1H3    ----------------------------............    90
LODC3'    ----------------------------.taagagaacaac    200

ODC1H1    ............................TAATAGAAC     99
ODC1H2    ............................---------     100
ODC1H3    ............................---------     99
LODC3'    ttactgatatatttatttgtagctaaaatca---------    240
```

Region 1 (contd.)

```
ODC1H1    AACTTACTGATATATTTATTT....GTAGCTAAAATGATA    135
ODC1H2    ----------t---------....t--a----------    136
ODC1H3    ----------------------....---------------    135
LODC3'    ----------------------attt---------------    280

ODC1H1    AACCAAAAGCTTTCCTTAACGCTGTTCCACTTATTTTTCA    175
ODC1H2    ----------------------------------------    176
ODC1H3    ----------------------------------------    175
LODC3'    ----------------------------------------g    320

ODC1H1    CTTCAGGCTGAACTTGGCTTCAATATGTATCTGCTTGACA    215
ODC1H2    ----------------------------------------    216
ODC1H3    ----------------------------------------    215
LODC3'    ----------------------------------------    360

ODC1H1    TTGGTGGTGGCTTCCCTGGCTCTGAAGACGTCAAGCTTAA    255
ODC1H2    ----------------------------------------    256
ODC1H3    ----------------------------------------    255
LODC3'    ----------------------------------------    400

ODC1H1    ATTTGAAGAGGTATAGCTTTGCTAAGAGCTTCGTGGTGTC    295
ODC1H2    ----------------------------------------    296
ODC1H3    ----------------------------------------    295
LODC3'    ----------------------------------------    440

ODC1H1    TTTCTGGTTGTTGAACTGTATTAAAATCTAGCTCTTCCCA    335
ODC1H2    -------a--------------------------------    336
ODC1H3    -------a----------------------------t---    335
LODC3'    -------a----------------------------t---    480

ODC1H1    TAGCATAAAGACACACTCATAAGAGTACATCTGATCAGAA    375
ODC1H2    ----------------------------------------    376
ODC1H3    ----------------------------------------    375
LODC3'    ----------------------------g-----------    520

ODC1H1    ATGGAATTTAGAA...AAGAGGAGACTGTATTTGGAAAAT    412
ODC1H2    ------------...------------------------    413
ODC1H3    ------------gaa------------------------    415
LODC3'    ------------gaa------------------------    560

ODC1H1    TAGATCATCTCGCAAGACAGAGCTAAATGAATGCATAAAC    452
ODC1H2    ----------a-----------------------------    453
ODC1H3    ----------t-----------------------------    455
LODC3'    ----------t----------------------------g    600

ODC1H1    ATGCAGAGATACCAAACTTGTAAAGCAAGTTTTAATTGCA    492
ODC1H2    ----------------------------------------    493
ODC1H3    ----------------------------------------    495
LODC3'    ----------------------------------------    640

ODC1H1    GACTTAATTCTCACTGGCATAAACTTCTTCCAGTTGACTG    532
ODC1H2    ----------------------------------------    533
ODC1H3    ----------------------------------------    535
LODC3'    ----------------------------------------    680
```

Region 1 (contd.)

```
ODC1H1    TCTACTTCAGGTGTTCTGAACTGATGCTGCTCCCACATAA   572
ODC1H2    ----------------------------------------   573
ODC1H3    ----------------------------------------   575
LODC3'    ---------t------------------------------   720

ODC1H1    GATGAGAAAACCCTGAACCCATACAGTCTTTGAGTGGTTT   612
ODC1H2    ----------------------------------------   613
ODC1H3    ----------------------------------------   615
LODC3'    ----------------------------------------   760

ODC1H1    TGAGCACTGGTGTAGCAAACCAGATCAGTAAGTGATCTGG   652
ODC1H2    ----------------------------------------   653
ODC1H3    -a--------------------------------------   655
LODC3'    -a--------------------------------------   800

ODC1H1    CTTATACATTCCTAGCACTAAGTATGGGTTTTTGTTCAAT   692
ODC1H2    ----------------------------------------   693
ODC1H3    ----------------------------------------   695
LODC3'    ----------------------------------------   840

ODC1H1    CAGTTCTCTAGTCTTTCATAGAAATACCTACTTTTCTGAA   732
ODC1H2    ----------------------------------.-a---   732
ODC1H3    ----------------------------------------   735
LODC3'    ----------------------------------.-----   879

ODC1H1    GAGGAGAGGTGGCTTAGTGTTAAGAGGACTAAAAGCAGTG   772
ODC1H2    -------------------------a------------a   772
ODC1H3    -------------------------a------------a   775
LODC3'    -------------------------a------------a   919

ODC1H1    TTGTCAGAACACCTGCCAGTCTAAAGTGATGTTAAATGGT   812
ODC1H2    ----------------------------------------   812
ODC1H3    ----------------------------------------   815
LODC3'    ----------------------------------------   959

ODC1H1    GTTATGTGCTTACGCAGGGAGACATCAAATCAGTCCCACA   852
ODC1H2    ----------------------------------------   852
ODC1H3    ----------------------------------------   855
LODC3'    ----------------------------------------   999

ODC1H1    GAGT.....................................   856
ODC1H2    ----.....................................   856
ODC1H3    ----.....................................   859
LODC3'    ----gaattatctgcttgtagtgtgggcttatgttgcctc  1039
```

(B) Sequence Alignment in Region 2

```
ODC2H1      ....................................TGCT      4
ODC2H2      ....................................----      4
ODC2H3      ....................................----      4
LODC3'      tgtactggctttatataaaatatcaagtgcaaatga----   1120

ODC2H1      GTGAGTGGGCTNNNTGGCGTGAGGCAGTCATGCAGTGGTT     44
ODC2H2      -------------------------------------------  44
ODC2H3      -------------------------------------------  44
LODC3'      ----------ccg-------------------------- 1160

ODC2H1      GGAGGCAGCTTTATTGCTGATAAGTTGGCAGCTGTGAAGG     84
ODC2H2      -------------------------------------------  84
ODC2H3      -------------------------------------------  84
LODC3'      ------------------------------------------- 1200

ODC2H1      CAACAATTTCCAAGATGATGACAAAGATTTCTGTAGTCTT     124
ODC2H2      -------------------------------------------  124
ODC2H3      -------------------------------------------  124
LODC3'      ------------------------c-------------- 1240

ODC2H1      AAAATCCCGTGAGATGACTTCTCTAGGATGCAGTAAACTG     164
ODC2H2      -------------------------------------------  164
ODC2H3      --------------------------c-------------  164
LODC3'      --------------------------c-------cc---- 1280

ODC2H1      GAGGGTGTTGCATTATGATCAGAATGAGAAGCATACCAGG     204
ODC2H2      -------------------------------------------  204
ODC2H3      -------------------------------------------  204
LODC3'      ------------------------------------------- 1320

ODC2H1      AATTGAAGTGGAAAGGTGTTGACACGTGGGACACTAATGC     244
ODC2H2      -------------------------------------------  244
ODC2H3      -------------------------------------------  244
LODC3'      ------------------------------------------- 1360

ODC2H1      TGTTTTAATGAAGCCTTTAGTGAAGGGTGACTTCACTGAG     284
ODC2H2      -------------------------------------------  284
ODC2H3      -------------------------------------------  284
LODC3'      ------------------------------------------- 1400

ODC2H1      ACTTAAGTGACCAAACAGCTCACGTTTAAGTTTTTAAGCT     324
ODC2H2      -------------------------------------------  324
ODC2H3      -------------------------------------------  324
LODC3'      ------------------------------------------- 1440

ODC2H1      GCAAGT..CTGCTTTTATTGTCTTCCACTTTGAAAATACA     362
ODC2H2      ------..-----------------------------------  362
ODC2H3      ------..-----------------------------------  362
LODC3'      ------gt---------------------------------- 1480

ODC2H1      ATAAAGTGAAGACAAATGATACAGCAAGCCACTTGTGTAT     402
ODC2H2      -------------------------------------------  402
ODC2H3      -------------------------------------------  402
LODC3'      ------------------------------------------- 1520
```

Region 2 (contd.)

```
ODC2H1    CCTAACCTGTGTGTTTGAATAGTACCTTACTGTGAGCATG    442
ODC2H2    ----------------------------------------    442
ODC2H3    ----------------------------------------    442
LODC3'    ----------------------------------------   1560

ODC2H1    GAAGTCTGACTTCATTCCTGTCTAAAAAGAGTATCTGGGT    482
ODC2H2    ----------------------------------------    482
ODC2H3    ----------------------------------------    482
LODC3'    ----------------------------------------   1600

ODC2H1    AGCTGCTGAAGTAGTAATATTTGCAACTGGTGTTACTTCC    522
ODC2H2    ----------------------------------------    522
ODC2H3    ----------------------------------------    522
LODC3'    ----------------------------------------   1640

ODC2H1    TAGATCACAAGTGTAATCAACCCAGCACTGGATAAATACT    562
ODC2H2    ----------------------------------------    562
ODC2H3    ----------------------------------------    562
LODC3'    ----------------------------------------   1680

ODC2H1    TTCCTTTGGATTCTGAAGTAACTATTATTGCAGAGCCAGG    602
ODC2H2    ----------------------------------------    602
ODC2H3    ----------------------------------------    602
LODC3'    ----------------------------------------   1720

ODC2H1    AAGATACTATGTTGCATCAGCATTCACCCTGGCAGTCAAT    642
ODC2H2    ----------------------------------------    642
ODC2H3    ----------------------------------------    642
LODC3'    ----------------------------------------   1760

ODC2H1    ATCATTGCAAAAAAGATTGTGTCAAAGGAGCAGACAGGTT    682
ODC2H2    ----------------------------------------    682
ODC2H3    ----------------------------------------    682
LODC3'    ----------------------------------------   1800

ODC2H1    CTGATGGTATGTGAATCTGGATTTGATCTTCCCTCCCCCA    722
ODC2H2    ----------------------------------------    722
ODC2H3    ----------------------------------------    722
LODC3'    ----------------------------------------   1840

ODC2H1    GCCTAACTGTATGT..........................    736
ODC2H2    --------------.........................    736
ODC2H3    --------------.........................    736
LODC3'    --------------ttctttccaagcactctcttactggt   1880
```

(C) Sequence Alignment in Region 3

```
ODC3H1      ..........................AACCAGTT      8
ODC3H2      ..............................--------      8
ODC3H3      ..............................--------      8
LODC3'      tcaattgcatctgtatgatcatgcacatgtta--------   2000

ODC3H1      CTGCAAAAGGTACAGCATTTCACAAGTGTTGCACGTTGGT     48
ODC3H2      ---------------a------------------------     48
ODC3H3      ------------------------------t-------      48
LODC3'      ------------------------------t-------   2040

ODC3H1      AGTTTGGCTGTAGGCAAATCCCTCCTCATAGCAGTCTCAT     88
ODC3H2      ---------------------------------------g-     88
ODC3H3      ---------------------------------------g-     88
LODC3'      ---------------------------------------g-   2080

ODC3H1      GTATTCACTGCTCCCTCTGCATAGCCAGAATCTGACATTA     128
ODC3H2      --..-------.....................---------     106
ODC3H3      --..-------.....................---------     106
LODC3'      --..-------.....................---------   2098

ODC3H1      AACTTGGGAGGCGTGCAGTGGAACTTAATTCAGAGCCAAG     168
ODC3H2      ----------------------------------------     146
ODC3H3      ----------------------------------------     146
LODC3'      -------.--------------------------------   2137

ODC3H1      GGTGCAGAAGGACCTGGGATCAGTTCTTCCTCCTCTTAGA     208
ODC3H2      ----------------ca--------...-----------     183
ODC3H3      ----------------ca--------...-----------     183
LODC3'      ----------------ca--------...-----------   2174

ODC3H1      GAAGAGGCAGAGTGCTCTGCATGCATCAATCATGACATAA     248
ODC3H2      --g----------------t--------------------     223
ODC3H3      --g----------------------------------------  223
LODC3'      --g----------------t---------------t----   2214

ODC3H1      CTTCTGGAAGCTGTTTTAGTTGTCAAACTAATGAGACTCT     288
ODC3H2      ----------------------------------------     263
ODC3H3      ----------------------------------------     263
LODC3'      ----------------------------------------   2254

ODC3H1      TCTCTTTAGCGGCCTAAACCAGATGACGGCTGCTACTCCT     328
ODC3H2      ----------------------------------------     303
ODC3H3      ----------------------------------------     303
LODC3'      ----------------------------------------   2294

ODC3H1      GCAGCATATGGGGACCAACGTGTGATGGCCTGGATCGTAT     368
ODC3H2      ----------------------------------------     343
ODC3H3      ----------------------------------------     343
LODC3'      ----------------------------------------   2334

ODC3H1      TGTTGAGCGTTGTAACATGCCAGAGTTGCAAGTTGGTGAC     408
ODC3H2      ----------------------------------------     383
ODC3H3      ----------------------------------------     383
LODC3'      ----------------------------------------   2374
```

Region 3 (contd.)

```
ODC3H1      TGGATCCTGTTTGAAAACATGGGTGCCTATACTGTTGCAG     442
ODC3H2      -----------------------------------------    417
ODC3H3      -----------------------------------------    417
LODC3'      -----------------------------------------   2414

ODC3H1      CAGCTTCTACTTTCAACGGATTCCAGAGGCCAACAATACA     482
ODC3H2      -----------------------------------------    457
ODC3H3      ----------------t------------------------    457
LODC3'      ----------------t------------------------   2454

ODC3H1      TTATGTGATGTCAAGACCAGCATGGTTAGTATCTTAAATA     522
ODC3H2      -----------------------------------------    497
ODC3H3      -----------------------------------------    497
LODC3'      -----------------------------------------   2494

ODC3H1      TGGTAATAACTTGGTCCACCAAGTGCTTTCATTAGAGAAA     562
ODC3H2      ----------------------------t----------    537
ODC3H3      -----------------------------------------    537
LODC3'      -----------------------------------------   2534

ODC3H1      CTGGTGAATATCTTCTCAGTTGTGACTGTCCTGTATCAAA     602
ODC3H2      ----------g----------------------c------    577
ODC3H3      ----------g----------------------c------    577
LODC3'      ----------g----------------------c------   2574

ODC3H1      CAGTGCCATGGCTGTGGTGTTGTATTGGCCGGAAAACCAT     642
ODC3H2      --------------a----c---------t----g-----    617
ODC3H3      --------------a----c---------t----g-----    617
LODC3'      --------------a----c--------------.-----   2613

ODC3H1      ATTTTTATGTGAAAAGCAACACCTCTGGTAGTCTTTGCAT     682
ODC3H2      -----------------------------------------    657
ODC3H3      -----------------------------------------    657
LODC3'      -----------------------------------------   2653

ODC3H1      GTTTCTAGAACAACCATAACACTACCAATTACTTTTTGCT      722
ODC3H2      -----------------------------------------    697
ODC3H3      -----------------------------------------    697
LODC3'      -----------------------------------------   2693

ODC3H1      CTGATTTATTTATCTCTAGAAGTGCTTTTTAAGAGAGAAT     762
ODC3H2      ----------------------------a----------    737
ODC3H3      ----------------------------a----------    737
LODC3'      ----------------------------a----------   2733

ODC3H1      GCT.........AGGAGTCCTTTCATTAGCCTGGGCTTCC     793
ODC3H2      ---.........-----------------------------    768
ODC3H3      ---ctttaaagaga--a------------------------    777
LODC3'      --g.........-----------------------------   2763

ODC3H1      AGCTTGTCTTTCACTTTTGTGAAAACATAAGGTAGTTACA     833
ODC3H2      -------------------c.--------------------    807
ODC3H3      -------------------c.--------------------    816
LODC3'      -------------------c.--------------------   2802
```

Region 3 (contd.)

```
ODC3H1      GATTCTTTACAGTTAACTTCTCTGACCAGCGATCAGAGAC    873
ODC3H2      -----------------------------a-------g-    847
ODC3H3      -----------------------------a-------g-    856
LODC3'      -----------------------------a---------   2842

ODC3H1      AGATCAATCACTGTTTTGCTATTGTGCTGTCTCGTAAGTG    913
ODC3H2      -----------c-----------a------t---------    887
ODC3H3      -----------------------------t---------    896
LODC3'      -----------c--------------------t-------   2882

ODC3H1      GATAGCCCGCTATATCCTGACTGTAAGGGACAGGAATGCA    953
ODC3H2      ----------------------------------------    927
ODC3H3      ----------------------------------------    936
LODC3'      -------t--------------------------------   2922

ODC3H1      AAGGAACAGTTCTTAACTGTGCAGCTCCACAATATGTTCT    993
ODC3H2      ----------------------------------------    967
ODC3H3      ----------------------------------------    976
LODC3'      ----------------------------------------   2962

ODC3H1      GTTTGAGAGCTGTTCGTGGG..........              1013
ODC3H2      --------------------.....                   987
ODC3H3      --------------------.....                   996
LODC3'      --------------------tgtaaattgagctccggact   3002
```

(D) Sequence Alignment in Region 4

```
ODC4H1      ..............................TCCGGA      6
ODC4H2      ..............................------      6
ODC4H3      ..............................------      6
LODC3'      ctgtttgagagctgttcgtgggtgtaaattgagc------   3000

ODC4H1      CTGTGGCTTACTGTAGTCTGTTGGATCTCCACAGGAAGCC     46
ODC4H2      -------------------------------------       46
ODC4H3      -------------------------------------       46
LODC3'      -------------------------------------    3040

ODC4H1      CTAAAATAGATGAGGAAAGCAAGGTTACTTT.AAGTACCA     85
ODC4H2      ------------------------------t--------     86
ODC4H3      ------------------------------.--------     85
LODC3'      ------------------------------.--------   3079

ODC4H1      GAATGCTGCAGAACTGCATCTTTAAACAGTAGCTAAAAGT    125
ODC4H2      --------------------------------t-------   126
ODC4H3      --------------------------------t-------   125
LODC3'      -------------------------------------    3119

ODC4H1      AATCTGGATGTTTTTCTCTGCAAAGCAGGTTAATGGACTT    165
ODC4H2      ------------------a------------------t--   166
ODC4H3      ------------------a------------------t--   165
LODC3'      ------------------a-------t----------t--   3159

ODC4H1      CTATAAAAAA.GGGTTCTTATGTAAGAGGAAGTCCTTATC    204
ODC4H2      ----------a--------------------------      206
ODC4H3      ----------a--------------------------      205
LODC3'      ----------a----------c-----------------   3199

ODC4H1      TTTACTGAAGCTTGCCTATTAAAATGGTAAAAAGTATTTG    244
ODC4H2      ----------t--------------------------      246
ODC4H3      ----------t--------------------------      245
LODC3'      ----------t--------------------------   3239

ODC4H1      GGCAGCAGTATTAGCATGTTTGGCTAGGATCAGGTGGTTC    284
ODC4H2      -------------------------------------       286
ODC4H3      -------------------------------------       285
LODC3'      -------------------------------------    3279

ODC4H1      TGAGATTTACTGTAAACTCTTAAATCTAGGGGAGCTCTGG    324
ODC4H2      -------------------------------------       326
ODC4H3      ----------------t----------------------     325
LODC3'      -------------------------------------    3319

ODC4H1      TTGTAAAAGAGGAAGCTTCCTCCATTCTTCCATTAGCCTT    364
ODC4H2      -------------------------------------       366
ODC4H3      -------------------------------------       365
LODC3'      -------------------------------------    3359

ODC4H1      GCCTTCCTATAAACAAGGTTGGCACTTCTGTTTGGCTTTT    404
ODC4H2      -------------------------------------       406
ODC4H3      -------------------------------------       405
LODC3'      -------------------------------------    3399
```

Region 4 (contd.)

```
ODC4H1      TCCCTCCCCTATCTTTAGTCACACCTCCAGTCAAATGAAG    444
ODC4H2      .--------c-c------cg--------------------    445
ODC4H3      .--------c-c------cg--------------------    444
LODC3'      .--------c-c------cg--------------------   3438

ODC4H1      TATTGC.TGTGTCACATGAAGCCTTAGGCTTCCCATTAGA    483
ODC4H2      ------.-----------------------------------  484
ODC4H3      ------.-----------------------------------  483
LODC3'      ------a-----------------------------------  3478

ODC4H1      CTCCAGACTTGTCTGAAACTAAAGGTGGCTCTGCAGAGTT    523
ODC4H2      -------------------------t----------------  524
ODC4H3      -----------------------------------------   523
LODC3'      ----------------------------------------c-  3518

ODC4H1      GTTAATTAACTTTTTCACTATAGACTTCTATTAAAATTGT    563
ODC4H2      -----------------------------------------   564
ODC4H3      -----------------------------------------   563
LODC3'      -----------------------------------------   3558

ODC4H1      AATATTTGAGCTGAGGCATACCATCCAAAGCTACAGTTCT    603
ODC4H2      -----------------------------------------   604
ODC4H3      -----------------------------------------   603
LODC3'      -----------------------------------------   3598

ODC4H1      TTTCACCCCTTCTTAGTGTTCTTCTTTTGGCAAGAAAATT    643
ODC4H2      -----------------------------------------   644
ODC4H3      -----------------------------------------   643
LODC3'      -----------------------------------------   3638

ODC4H1      CTACCAGCATATGCTAACCCTTGTCTGAGGGGGATATAAT    683
ODC4H2      -----------------------------------------   684
ODC4H3      -----------------------------------------   683
LODC3'      -----------------------------------------   3678

ODC4H1      CAGTGAGCTTATGCTGCATGTCCTTTCAGCACAGACAAAA    723
ODC4H2      -----------------------------------------   724
ODC4H3      -----------------------------------------   723
LODC3'      -------g---------------------------------   3718

ODC4H1      GCTTTTTTGATACGTGCTCTACGTGTTTAAGTGTACAGCT    763
ODC4H2      -----------------------------------------   764
ODC4H3      -----------------------------------------   763
LODC3'      ------------t----------------------------   3758

ODC4H1      GTATTTATAGCTCTAACTGTGATACGTGTAAGGGATTACC    803
ODC4H2      -----------------------------------------   804
ODC4H3      -----------------------------------------   803
LODC3'      -------------....-------t---------------   3794

ODC4H1      TATGTCTGAGTGCATATCACATGGGCAACCATGTGGTAAA    843
ODC4H2      -----------------------------------------   844
ODC4H3      -----------------------------------------   843
LODC3'      -----------------------------------------   3834
```

Region 4 (contd.)

```
ODC4H1      CACAACTTTTCTTTGCATGCAGTGTAAGATACGTAGTCAA    883
ODC4H2      ----------------------------------------    884
ODC4H3      ----------------------------------------    883
LODC3'      ----------------------------------------   3874

ODC4H1      AAACTAATTAACTGAAGTTCTTCTGTACTGAGCTTGTTTT    923
ODC4H2      ----------------------------------------    924
ODC4H3      ----------------------------------------    923
LODC3'      ------------a---------------------------   3914

ODC4H1      GGTACAGTGTTTGAGGATGAAGGGGAGTTGTGGTTCTCCT    963
ODC4H2      ----------------------------------------    964
ODC4H3      ----------------------------------------    963
LODC3'      ----------------------..----------------   3952

ODC4H1      ATGCCAACTCCCTTAAATTCCTGCTGT.............    991
ODC4H2      ---------------------------.............    992
ODC4H3      ---------------------------.............    991
LODC3'      g--------------------------taaccgcacttga   3992
```

(D) Sequence Alignment in Region 5F

```
Odc5FH1     ......................TCATGAAACCTGTATGG     17
Odc5FH2     ......................---c-------------     17
Odc5FH3     ......................---c-------------     17
LODC3'      ttcattctggaatgacttctaag-------/---------   4080

Odc5FH1     CATGTGGAATAAAAATGTAGGTCCTGATATATGCATATCT     57
Odc5FH2     --c---------------------t-------------     57
Odc5FH3     --c---------------------t-------------     57
LODC3'      --c--------------------t---------t------   4120

Odc5FH1     GCAGATACAAATCCTACTTGGAAGCAGTGAAACAGCTTAA     97
Odc5FH2     ---------g-------------------------------    97
Odc5FH3     ---------g-------------------------------    97
LODC3'      -------g-g-------------------------------   4160

Odc5FH1     TGTTGGCATTTACACTAAAGACTGCAGATGACATGCACTG     137
Odc5FH2     -a-------------------------------------a-    137
Odc5FH3     -a-------------------------------------     137
LODC3'      -a--------------------c---------------a-   4200

Odc5FH1     ACTGTTCTGCTTTGGTATTTTGTTCTAGTAATAATGTCTG     177
Odc5FH2     ------------------------------------a----    177
Odc5FH3     ----------------------------------------     177
LODC3'      ------------------------------------a----   4240

Odc5FH1     GATTGTAAAGCATCGATTNNNNNNNNNNNNNTGGTAGCTAC    217
Odc5FH2     ----------------------------------------     217
Odc5FH3     ----------------------------------------     217
LODC3'      ------------------ttttttgtgttc----------   4280

Odc5FH1     AAAGTGGTAACTAAACTATTTCATCTATCTTTTCTAGGAT     257
Odc5FH2     gt---a-----------------------------------    257
Odc5FH3     ----------------------------------------     257
LODC3'      ----------------------------------------   4320

Odc5FH1     GTTGTATCTGAAATACTCTTCTGACTGTTGGTGTTTTTCT     297
Odc5FH2     ----------------------------------------     297
Odc5FH3     ----------------------------------------     297
LODC3'      -------.--------------------------------   4359

Odc5FH1     TTCTAGGCAATTAATGCAACAGATAAAGGAGCAGGAGTTC     337
Odc5FH2     ----------------------------------------     337
Odc5FH3     ----------------------------------------     337
LODC3'      ----------------------------------------   4399

Odc5FH1     CTAGCTGAAGTGG..........................     350
Odc5FH2     -------------..........................     350
Odc5FH3     -------------..........................     350
LODC3'      -------------aggagcaggatgttgctagtctgcccc   4439
```