# Genetic control and variation in turkey: molecular insights in selection

**Muhammad Luqman Aslam**

**Thesis committee**

**Thesis supervisor**
Prof. dr. M.A.M. Groenen
Personal chair at the Animal Breeding and Genomics Centre
Wageningen University

**Thesis co-supervisors**
Dr. ir. J.W.M. Bastiaansen
Research Associate at the Animal Breeding and Genomics Centre
Wageningen University

Dr. R.P.M.A. Crooijmans
Assistant professor at the Animal Breeding and Genomics Centre
Wageningen University

**Other members**
Prof. dr. ir. J. Keijer, Wageningen University
Dr. ir. M.C.A.M. Bink, Wageningen UR
Prof. dr. D.K. Griffin, University of Kent, School of Biosciences, United Kingdom
Dr. P. van As, Hendrix Genetics, Research and Technology Centre, Boxmeer

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

# Genetic control and variation in turkey: molecular insights in selection

**Muhammad Luqman Aslam**

**Thesis**
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis committee appointed by the Academic Board
to be defended in public
on Monday 17 September 2012
at 4 p.m. in the Aula.

*Dedicated to my beloved parents, sisters, brothers and wife*

## Abstract

Aslam, M.L. (2012). Genetic control and variation in turkey: molecular insights in selection. PhD thesis, Wageningen University, the Netherlands

The turkey (*Meleagris gallopavo*) is an important agricultural species that is largely used as a meat type bird as egg production of this species is very low. Turkey is the second largest contributor to the world's poultry meat production after chicken. Understanding the etiology and biology underlying production and health traits is very important for the genetic improvement of these traits in the desired direction and to avoid undesired side-effects. The aim of the research described in this thesis was to interrogate the genetics of turkey traits related to meat production and to investigate the genetic diversity of commercial and heritage turkey populations. Different analyses were performed that included the estimation of genetic and (common) environmental variances for growth (body weight as well as growth curve traits), breast meat yield and meat quality traits in turkeys. I describe the construction of a single nucleotide polymorphism (SNP) based linkage map of turkey and its comparison with the physical map of chicken to investigate genome structural differences between these highly important poultry production species. Two inter-, and 57 intra-chromosomal rearrangements between these two species were confirmed or discovered which is a low number in comparison to mammals and lead to the conclusion that turkey and chicken have highly conserved genomic structure. I used the linkage map of turkey together with individual phenotypes to map quantitative trait loci (QTL) in the same population for the traits described above. Results showed quantitative trait loci on 21 of the 27 turkey chromosomes covered by the linkage map. Forty-five quantitative trait loci were detected across all traits and these were found in 29 different regions on the 21 chromosomes. The next step, after the analyses on the reference population was to investigate the genomic variation in turkeys Next generation sequencing was used to investigate genome variation and the discovery of genome-wide signatures of selection in the turkey respectively. Sequencing was performed on 32 individuals from eleven different turkey populations (seven commercial, three heritage and a South Mexican wild population). Analysis of next generation sequencing data resulted in the detection of 5.49 million putative SNPs compared to the reference genome. The average frequency of heterozygous nucleotide positions in individual turkeys was 1.07 Kb-1 which is substantially lower than in chicken and pigs. The SNPs were subsequently used for the analysis of genetic diversity between the different populations. Genetic diversity analysis using pairwise Nei's genetic distance among all the individuals from the 11 turkey populations showed that all of the

commercial lines branched from a single node relative to the heritage varieties and the ancestral turkey population, indicating that commercial lines appear to share a common origin.

After assessing genome wide variation and diversity between breeds, the SNP data from ten of the turkey populations (29 individuals) was used to detect selective sweep regions. Across the turkey populations, 54 genomic regions with significant evidence for a selective sweep were detected. These sweeps were distributed over 14 different chromosomes. This study has investigated the genetics i.e. analysis of variances and QTL mapping related to economically important traits in turkey production and the genomic variation of turkey. Furthermore, this study has also created resources e.g. millions of discovered SNPs for subsequent genomic work in the turkey such as to discover variant (s) for both minor and major effects on traits of economic importance, and a high-resolution linkage map can be developed.

# Contents

# 1

## General Introduction

## 1.1 Introduction

### Taxonomy of turkey

Turkeys are classified in the taxonomic order of Galliformes of the genus *Meleagris*. Within this order they are relatives of the family/subfamily Tetraonidae (grouse) [1, 2]. There are still two living species occurring in the wild, the *Meleagris gallopavo* and the *Meleagris ocellata* . The *Meleagris gallopavo*, commonly known as the Wild Turkey, is native to the forests of North America [3] and the *Meleagris ocellata* or *Ocellated* Turkey, is native to the forests of the Yucatán Peninsula, Mexico [1, 2]. There are several extinct species, with archeological samples dating from as far back as 23 million years ago [4].

Turkeys do have characteristic features such as a distinctive fleshy wattle that hangs from the underside of the beak, and a fleshy protuberance that hangs from the top of its beak called a snood. With wingspans of 1.5–1.8 metres (4.9–5.9 ft), the turkey is by far the largest bird in the open forests in which they live. As in many galliform species, the female (hen) is smaller and if breed/species have a colored feather phenotype (not white) the hen is less colorful than the male (tom or gobbler).

### Habitat of wild turkeys

The natural habitat used by wild turkeys varies considerably according to the season, climatic conditions and performed behavior. Turkeys regularly utilize environments as diverse as open plains, dense woodland, thick scrub, and treetops, and can sometimes even be seen wading in lakes. The walking speed of the wild turkey is approximately 5 km/h but birds can run with great maneuverability at speeds of up to 30 km/h. Although their endurance is not great, wild turkeys are capable of flight in contrast to the domesticated strains [5]. Wild turkeys are not true migrants but can move up to 80 km between winter and summer sites. Typically, daily movement is 2±3 km and the home range covers from 0.81 to 4.04 $Km^2$ [6, 7].

### History and domestication of turkeys

The domestic turkey is derived from the native wild turkey of North America. There are seven subspecies of the wild *Meleagris gallopavo* [8] distinguished by geographic range and plumage differences. The subspecies are: Mexican (*M. g. gallopavo*), Rio Grande (*M. g. intermedia*), Merriam's (*M. g. Merriami*), Gould's (*M. g. mexicana*), Eastern (*M. g. silverstris*), Moore's (*M. g. oneusta*) and Florida (*M. g. osceola*). Three of these seven are important in the history of turkey domestication. It is generally accepted that the first ancestor of the domestic turkey was the *Meleagris gallopavo gallopavo* [9]. The Eastern wild turkey, *M. g. silverstris,* later hybridized with Mexican domesticated turkeys to form the commercial turkey. The

13

Merriam's wild turkey, *M. g. Merriami,* was domesticated separately in the south western part of current USA but has neither persisted nor contributed to present day commercial turkeys.

Turkeys taken to Europe during the 1500s were descendants of the Mexican turkey but since these original transportations occurred. There are a total of five wild turkey subspecies in North America that are not genetically related to modern commercial lines. Perhaps the most plausible derivation of the popular name is that when these birds were first introduced to Europe, anything foreign was said to be from Turkey and this word eventually became associated with the species [6, 7].

During the short time since their domestication, modern strains of turkeys have been selected for high growth rate and increased mature body size. Mature toms are too large to achieve natural fertilization without injuring the hens, so their semen is collected, and fertilization performed by artificial insemination (AI) to protect the welfare of the female bird. Although domestic birds retain many of the characteristics of their wild relatives, there are also fundamental differences. The vast majority of domestic turkeys are from a very small number of strains, most of which have completely white plumage, though some have retained the wild type mottled appearance. Commercially, male turkeys are routinely grown to approximately 20 weeks of age with a weigh of over 20 kg, this in contrast to the 9 kg of a 3-year-old male wild turkey [10]). Perhaps the most obvious difference in behaviour between the wild and domestic turkey is the inability of the latter to fly probably because of change in body texture through time. Domestic turkeys have retained the ability to run quickly, especially at younger ages.

### *Recognized turkey varieties*

There are currently eight varieties (Figure 1.1) of domestic turkeys recognized by the American poultry Association [11]:

| | |
|---|---|
| 1) Royal Palm Turkey (RP) | 2) Blue Slate Turkey (BLS) |
| 3) Beltsville small White Turkey (BvSW) | 4) White Holland Turkey (WH) |
| 5) Narragansett Turkey (Nset) | 6) Bronze (Bz) |
| 7) Black Spanish Turkey (BL) | 8) Bourbon (Bo) |

In addition to the recognized varieties, many more exist as officially unrecognized variants or as recognized breeds in other countries :

**Auburn**, an extremely rare heritage variety, numbers are not considered high enough for inclusion in the Standard.

**Broad Breasted White,** a non-standardized commercial strain that does not qualify as a breed, only used for commercial meat production.

**Broad Breasted Bronze,** a non-standardized commercial strain that does not qualify as a breed, only used for commercial meat production.

14

**Bronze Turkey**, the heritage strain of the Bronze is recognized, while the **Broad Breasted Bronze**, like the Broad Breasted white, is an unrecognized commercial meat strain.

**Chocolate**, Chocolate Brown in color. Day-old poults are white faced with chocolate bodies.

**Midget White**, a rare heritage variety sometimes conflated with the Beltsville Small White.

### *Turkey production*

The turkey (*Meleagris gallopavo*, MGA) is an important agricultural species that is largely used as a meat type bird as egg production of this species is very low. Turkey is the second largest contributor to the world's poultry meat production after chicken. The turkey bird is easy to raise, does not require any special attention, is hardy and is less prone to diseases as compared to chicken [12]. Turkey is completely resistant to Marek's disease and I.B [12].

In 2009, turkey represented 5.8% of the world poultry meat production[13]. The world-wide turkey population has rapidly grown due to increased commercial farming. Global turkey stocks tripled from 178 million in 1970 to over 548 million in 2009. Over the same time period, the production volume increased more than fivefold from 1.2 to 5.3 million tons[13].

### *Genetic nature of phenotypic traits*

Genetically, traits can be divided into single gene controlled traits (monogenic traits) and the traits which are controlled by a number of genes (polygenic or quantitative traits). Monogenic traits follow the pattern of Mendelian inheritance while polygenic or quantitative traits don't follow the pattern of Mendelian inheritance. The example of monogenic traits in birds are plumage color, fishy taint in eggs, naked neck etc. while polygenic traits include most of the economically important traits e.g. growth traits, reproduction related traits, meat quality traits etc. The detection of the causative variant for the monogenic traits is relatively easy and more prone to success with examples in different livestock species [14-17]. Whereas, the detection of causative variants for polygenic or quantitative traits is a very complex process because these traits are controlled by more than one locus and also the environment can have an effect on the trait phenotype [18, 19]. Accurate detection of causative variant/QTL regions for the polygenic traits requires highly sophisticated genomic resources (whole genome mapping, next generation sequencing etc.) as well as unbiased trait phenotype recording [20, 21].
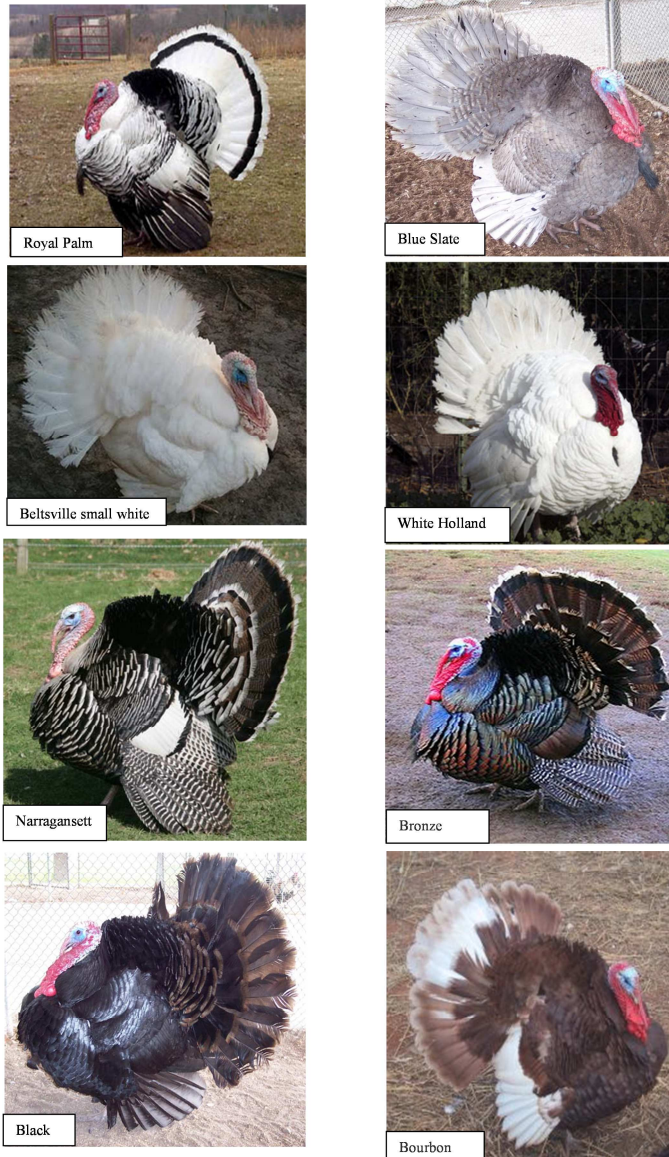
**Figure 1.1** Officially recognized turkey heritage varieties. The pictures were downloaded from the web sites. (http://poultrykeeper.com/narragansett-turkeys/the-narragansett-turkey/narragansett-turkey-photos.html; http://www.albc-usa.org/cpl/wholland.html; http://www.cacklehatchery.com/turkeypage.html; http://maryeaudet.hubpages.com/hub/Raising_Turkeys_for_Food_an_Profit_on_the_Home stead

### Variance components and heritability of polygenic traits

The phenotypic value for a specific individual is the result of genetic factors, environmental factors, and the environmental factors that interact with the genetic factors. The sum of these factors in a population segregating for a quantitative trait contributes to the variance of that population. Thus the total variance consists of the following components.

$$V_P = V_G + V_E + V_{GE}$$

$V_P$ = total phenotypic variation; $V_G$ = genetic variation that contributes to the total phenotypic variation; $V_E$ = environmental contribution to the total phenotypic variation; $V_{GE}$ = variation associated with the genetic and environmental factor interactions.

The genetic variation can be further subdivided into three components.

$$V_G = V_A + V_D + V_I$$

$V_G$ = total genetic variation; $V_A$ = additive genetic variance; $V_D$ = dominance genetic variance

$V_I$ = interaction genetic variance

and the total phenotypic variance can be rewritten as

$$V_P = V_A + V_D + V_I + V_E + V_{GE}$$

By performing specific experiments genetic and the environmental variances can be separated from the total phenotypic variance.

Heritability is an important genetic parameter which gives the information about the portion of additive variance which will be contributed from parents to their offspring. In general heritability can be described as the proportion of the genetic variance to the total variance. Heritability can be estimated in two ways. The broad-sense heritability is the ratio of total genetic variance to total phenotypic variance.

$$H^2 = V_G/V_P$$

The narrow-sense heritability is the ratio of additive genetic variance to the total phenotypic variance.

$$h^2 = V_A/V_P$$

### Estimation of variance components

For any trait of interest, observed differences among individuals may be due to differences in the genes coding for this trait or may be the result of variation in environmental condition. In many cases it is a combination of the two. Understanding the amount that genes, passed from parent to offspring, influence a trait may be useful in a variety of situations e.g. it can be useful to know what can be mature weight of the offspring from a particular cross or it can be useful in determining an individual's risk of developing a specific disease. The estimation of

17

genetic parameters is an important issue in animal breeding. First of all, estimating additive genetic and possible non-additive genetic variances contributes to a better understanding of the genetic mechanism. Secondly, estimates of genetic and phenotypic variances and covariances are essential for the prediction of breeding values (selection index and BLUP) and for the prediction of the expected genetic response of selection programs [22, 23]. Parameters that are of interest are heritability, genetic and phenotypic correlation and repeatability, and those are computed as functions of the variance components.

### *Selective breeding*

Historically, quantitative genetics-based selection has been the primary strategy of genetic improvement of livestock [24]. This genetic improvement is attributed largely to selective breeding programs that rely on highly heritable phenotypic traits, such as body size and breast muscle development. The efficiency of these classical methods used for genetic improvement decreases when traits are difficult to measure or have a low heritability [24]. The availability of genome-based selection using a large number of markers has the power to transform the breeder operation and incorporate previously unavailable genetic information into commercial lines [25].

Breeding programs for meat type birds are commonly selecting for body weight and body composition traits (breast yield, etc.) while minimizing production costs. Recently breeders have started to measure meat quality (drip loss, pH, etc.) as well as survival traits, at least in research projects [22, 26]. A number of publications are available for the estimation of genetic parameters for different traits in birds and the many other livestock species [27-30].

Selective breeding can be regarded as a long-term human experiment to alter the phenotypes of domesticated species. This kind of human experiments are expected to leave a signature in the genome of domesticated species [31, 32], for instance unusually low nucleotide diversity [31, 32] or the presence of exceptionally long haplotypes [33, 34]. Genome wide characterization for many different breeds and populations for these signatures of selection along with the functional knowledge of the region can reveal which genes are linked to traits or diseases with a complex genetic basis [35]. The genetic variation in domesticated species can thus be highly useful not only to gain a better understanding of consequences of selective breeding, but can also aid in elucidating fundamental biological and molecular pathways.

### *Genetic markers*

Genetic markers can be described as an observable variation in the DNA sequence which may arise due to mutation or alteration in the genomic loci. The variation

18

can be either one base alterations (single nucleotide polymorphism, SNP) or multiple bases such as variation in short or variable number tandem repeats. Commonly used genetic markers in molecular genetics based on single base variations are single nucleotide polymorphism (SNP), restriction fragment length polymorphism (RFLP), simple sequence length polymorphism (SSLP), amplified fragment length polymorphism (AFLP) and random amplification of polymorphic DNA (RAPD). Variation in repeat length are measured as variable number tandem repeat (VNTR) or simple sequence repeat (SSR).

SNP markers are momentarily the most used type of markers in genetic studies, and SNP markers can be found in high abundance within the genome [36]. SNP based genotyping is preferred because of it is high accuracy, quick and easily automated and using limited human intervention. SNPs are evolutionary stable, not changing significantly from generation to generation. This low mutation rate makes SNPs excellent markers for studying complex genetic traits and as a tool for understanding genome evolution [37]. Increasing the marker density of the linkage map further enables the analyses of genomic sequences associated with high recombination rates [38]. SNP markers can be rapidly and cheaply identified using DNA sequence data through different alignment or bioinformatics approaches [39-41]. SNP markers can be utilized to explore many aspects related to genetics, such as the detection of associations with certain traits of interest, genetic diversity studies, paternity assessment, forensics and inferences of population history [42, 43].

### *Genetic linkage mapping*

A genetic linkage map of a species or experimental population shows the order and distance of its genetic markers or known genes relative to each other in terms of recombination frequency (centimorgan; cM), rather than as specific physical distance (basepair; bp) along each chromosome. Molecular markers have revolutionized genome mapping over the last three decades, offering the potential for generating very high density genetic maps that can be used to develop haplotypes for genes or regions of interest, and whole genome mapping became a reality [44-53]. A genetic map represents the linear arrangement of markers on a chromosome and maps are prepared by analyzing populations derived from crosses of genetically diverse parents, and estimating the recombination frequency between genetic loci. The utilization of common molecular genetic markers across related species permits the comparison of linkage maps [54, 55]. This allows the translation of information between model species with sequenced genomes and non-model species [56]. Physical maps are based on the direct analysis of DNA sequence. Physical distances between and within loci are measured in basepairs

(bp), kilobasepairs (Kb) or Megabasepairs (Mb). Linkage and physical maps should provide the same information on chromosomal assignment and the order of loci. However, the relative distances that are measured within each map can be quite different [47, 53, 57]. Physical maps are high resolution maps and can provide an accurate description of the actual length of DNA that separates loci from each other. Linkage distances (cM) among the loci can be translated into estimated physical distances (bp or Mb) e.g. in chicken the genome average recombination rate estimated from the linkage map amounts to 3.1 cM/Mb [53]. Comparisons between genetic and physical maps clearly show that the rates of recombination vary considerably between species [58] and even between different chromosomes within a species [59]. The recombination rate in chicken is almost two-fold higher than in humans, where the recombination rate is about 1.2 cM/Mb [60] and two-fold higher than estimates from the zebra finch [52]. Even lower rates of recombination have been reported for rodents (rat and mouse 0.5 cM/Mb [58]. Among vertebrates, birds have a relatively high rate of recombination [41].

At the start of the study described in this thesis, limited information was available on the turkey linkage and physical map although a small number of low resolution linkage maps using microsatellite markers [61, 62] had been published. Comparative cytogenetic and linkage maps between turkey and chicken showed conserved synteny and close ancestral relation among these species [61, 62] and support the hypothetical ancestral Galliform karyotype [63].

### *QTL mapping*

Quantitative trait loci (QTL) are genomic regions with genes that directly or indirectly affect a quantitative trait [64]. Mapping those regions of the genome that contain genes which effect a quantitative trait is done using molecular markers (SSR, SNP, AFLP etc.) that are associated with the recorded phenotypes on a sample population. QTL mapping can be an early step in identifying and sequencing the actual genes underlying the causative mutation. A linkage map is essential for the mapping of QTL and very useful for the assembly of genome sequences and subsequently mapping of genes along the chromosomes. A high-resolution linkage map facilitates fine mapping of QTL and can be produced because of the abundance of SNPs within the genome [36].

Several studies have indicated that knowledge about genetic markers linked to genes affecting quantitative traits can increase the selection response of animal breeding programs, especially for traits that are difficult to improve by traditional selection [65, 66]. A large number of studies are available on QTL mapping for the growth, meat quality and the body composition traits of chicken [67-71] showing significant effects of QTLs on these traits of economic importance in poultry

20

breeding. Several studies reported significant association between individual genetic markers and quantitative traits of economic importance in chicken [67, 72-74] but no such reports exist for turkey.

*Next generation sequencing*

Next Generation Sequencing (NGS) or massively parallel sequencing (MPS) refers to a group of new DNA sequencing technologies that can rapidly sequence DNA on the gigabase scale. These methods have replaced classical first generation Sanger sequencing [75], which was the dominant sequencing technology from the late 1970's to the late 2000's and was used for all of the initial genome sequencing projects (H. influenzae, yeast, Drosophila, Arabidopsis, Human, Chicken etc.). The major novel advances offered by NGS are the ability to produce an enormous volume of sequence data cheaply and at high speed, in some cases in excess of one billion short reads (36-400 bp) per instrument run [76, 77]. Due to the variety of NGS features multiple platforms coexist in the marketplace, with some having clear advantages for particular applications over others e.g. Roche 454 Life Sciences, Illumina, Life Technologies SOLiD and Helicos Biosciences. The Roche 454 platform generates longer sequences (200-500 bp or more, depending on the version of the platform) than Illumina (35-150 bp) or SOLiD (25-75 bp), but SOLiD and Illumina have higher throughputs than Roche 454 with the same cost and time investment [39, 77].

The availability of a high quality reference genome sequence and the resequencing of individuals with appropriate coverage (multiple of copies from the whole genome) are essentials for the identification of genome-wide sequence differences. Sequence differences can be used, either for evolutionary studies or for discovering genetic variation that may explain phenotypic variation [39, 40, 78, 79]. For the discovery of genome variations among different individuals from the same or different species short sequence reads are mapped to the reference genome using different tools [80]. For the accurate and efficient mapping of these short sequence reads to the reference genome requires filtering control steps e.g. max read depth and mismatch percentage [40, 79, 81].

*Evolution of avian genomes*

Genome variation provides the necessary raw material for evolution by natural selection. In terms of the appearance of new variants, genomic variations are usually thought of as point mutations or short insertions and/or deletions in protein-coding or regulatory sequences, potentially resulting in phenotypic changes [82, 83]. Genome comparison of individuals of the same species or of different species can help in getting information about signatures of selection and to understand the function and evolutionary processes that act on genomes [31, 32,

21

62, 84-88]. Comparative genomics exploits both similarities and differences in the proteins, RNA, and regulatory regions of different organisms providing insight on how selection has acted upon these elements [32, 89, 90]. Those elements that are responsible for similarities between different species should be conserved through time (purifying or stabilizing selection) [91], while those elements responsible for differences among species should be divergent (positive selection) [92, 93]. Finally, those elements that are unimportant to the evolutionary success of the organism will be unconserved (selection is neutral) [93].

Chromosome studies have revealed that the karyotype is more conserved among avian lineages than it is among other groups, such as mammals, with most avian species showing a diploid chromosome number between 76 and 80 (http://www.genomesize.com). This suggests that chromosomal evolution or large-scale rearrangements affecting chromosome number occur at a low rate in birds and as a result many chromosomes have remained more or less intact during avian evolution [86]. Compared with many other organisms, avian karyotypes comprise chromosomes that differ significantly in size (Figure 1.2). Smaller chromosomes are often referred to as micro-chromosomes while larger chromosomes (comparable in size to typical mammalian chromosomes) are called macro-chromosomes. However, the size distribution of chromosomes is often continuous rather than bimodal and therefore the definition of macro- and micro-chromosomes is therefore somewhat arbitrary. The turkey genome consists of 39 pairs of autosomes and 1 pair of sex chromosomes [62]. The predicted size of the turkey genome is 1.1 billion bases [94].
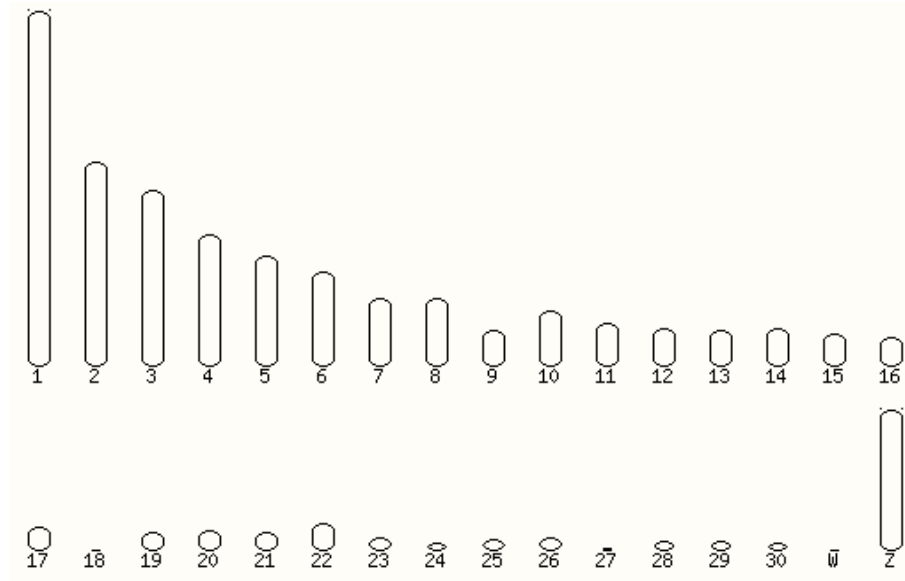
**Figure 1.2** Karyotype of turkey chromosomes showing larger as well as smaller chromosomes along with sex chromosomes (W and Z). (www.ensembl.org).

## 1.2 Aim and outline of thesis

The research described in this thesis aimed to (i) identify and investigate genetic control (chapter 2), (ii) identify causative variants (genomic regions) underlying variety of traits (chapter 4), (iii) map genomic regions that are or have been under selection during domestication and breeding (chapter 5 & 6) and to improve and increase available genomic resources in turkey.

We had access to a turkey population that was based on parents from two different lines that were crossed to produce full-sib families in the F1 generation. An F2 generation of 18 full sib families was produced by crossing 17 randomly selected F1 males and 18 randomly selected F1 females. Several phenotypic traits were measured and recorded on individuals of the F2 generation. These recorded traits were first used to describe variance (chapter 2) under genetic control for the variety of different traits. We estimated genetic parameters (heritability, genetic and phenotypic correlations) for different growth (body weight and growth curve traits), breast meat yield and meat quality traits in turkeys. Estimates of heritability, genetic and phenotypic correlations among different traits are very important when considering multiple trait improvement and selection. Heritability of a particular trait gives an idea whether it can be improved or not or how faster it will

be improved while correlation among traits describes the direction of a correlated trait(s) if selection is done on a desired trait(s). Chapter 3 describes the construction of the SNP based linkage map of turkey and its comparison with the physical map of chicken to investigate structural differences between the genomes of these highly important poultry species. The aims of this study were to improve and increase the available turkey genomic resources, to assist in the assembly of the turkey genome sequence and to use this linkage map for subsequent QTL mapping study. The next phase was to combine the obtained phenotypes (chapter 2) and genotypes (chapter 3) for the identification of QTL (chapter 4). We used the available phenotypic data and our linkage map (chapter 2 & 3) of the turkey to map QTLs for different traits such as; growth curve, body weight, breast yield and meat quality traits. In this chapter, we also compared the location of the quantitative trait loci identified in turkey, with the syntenic regions in chicken. In chapters 5 and 6, we describe the use of next generation sequencing to investigate genome variation and genome-wide signatures of selection during domestication and breeding in the turkey respectively. In chapter 5 we describe the discovery of 5.49 million putative SNPs that represents a powerful resource for subsequent genomic work in the turkey and for the development of a high-density SNP chip. These SNPs were subsequently used for the analysis of genetic diversity among the different populations. The same SNPs were later also used for a selective sweep study (chapter 6), using the twenty nine sequenced individuals of the ten different turkey populations that are described in chapter 5. Genome-wide signatures of selection or domestication (selective sweep regions) were identified based on the distribution of the heterozygosity pattern in the genome. The identified sweep regions were subsequently examined for the presence of QTL and the sweeps within the syntenic regions of chicken. In the general discussion (chapter 7) the findings presented in this thesis are discussed, in relation to the role of specific genes in controlling complex economically important traits and ways/techniques to identify variants in the genome that may affect the performance of individuals for a specific trait. Furthermore, this chapter discusses current selection procedures applied in the turkey industry and how genomic variants can be used effectively by the breeding industry.

# References

1.    Banks RC, McDiarmid RW, Gardner AL: **Checklist of Vertebrates of the United States, the U.S. Territories, and Canada**. Washington, D.C., USA: United States Department of the Interior Fish and Wildlife Service 1987.

2.    Dickinson EC: **The Howard and Moore Complete Checklist of the Birds of the World**, 3rd edn. Princeton, New Jersey, USA: Princeton University Press; 2003.

3.    Union AO: **The A.O.U. check-list of North American birds**. In*.*, 7th edn; 2004.

4.    D. S. Farner, J. R. King: **Avian biology**. In*.* Boston: Boston Academic Press; 1971.

5.    Eaton S: **The Birds of North America**. Philadelphia, PA: The Academy of Natural Sciences; Washington, DC: The American Ornithologists' Union.; 1992.

6.    Schorger AW: **The wild turkey: Its history and domestication**, 1st edn: University of Oklahoma Press(Norman); 1966.

7.    R. D. Crawford: **Evolution of domestic Animal : Chapter 47. Turkey**. New York; 1984.

8.    R. Howard, A. Moore: **A complete checklist of birds of the world**, Revised edn. London: Macmillan; 1984.

9.    Crawford RD: **Introduction to Europe and diffusion of domesticated turkeys from the America**. *Arch Zootec* 1992, **41** (extra):307-314.

10.   L. Williams: **The book of the wild turkey**. Tulsa: Winchester Press; 1981.

11.   Frank R, Reese J, Bender M, Sponenberg DP, Williamson D, Beranger J: **Selecting your best turkeys for breeding**. In*.*: The American Livestock Breeds Conservancy.

12.   TNAU Agritech Portal: **Animal Husbandry**. In*.* Coimbatore: TamilNadu Agriculture University; 2011.

13.   **Food and agriculture organization statistical division (FAOSTAT) of the United Nations** [http://faostat.fao.org/]

14.   Seitz JJ, Schmutz SM, Thue TD, Buchanan FC: **A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle**. *Mamm Genome* 1999, **10**(7):710-712.

15.   Tobita-Teramoto T, Jang GY, Kino K, Salter DW, Brumbaugh J, Akiyama T: **Autosomal albino chicken mutation (ca/ca) deletes hexanucleotide (−ΔGACTGG817) at a copper-binding site of the tyrosinase gene**. *Poult Sci* 2000 **79**:46-50.

16.     Mariat D, Taourit S, Guerin G: **A mutation in the MATP gene causes the cream coat colour in the horse**. *Genet Sel Evol* 2003, **35**:119-133.

17.     Gunnarsson H, Hellström AR, Tixier-Boichard M, Minvielle F, Bed'hom B, Ito S, Jensen P, Rattink A, Vereijken A, Andersson L: **Mutations in SLC45A2 cause plumage color variation in chicken and japanese quail**. *Genetics* 2007, **175**(2):867-877.

18.     Ulizzi L, Correani R, Terrenato L: **Effects of environmental changes on the phenotypic expression of human polygenic traits**. *Ann Hum Genet* 1988, **52**(4):341-349.

19.     Andersson L: **Genetic dissection of phenotypic diversity in farm animals**. *Nature Rev Genet* 2001, **2**:130-138.

20.     Dekkers JCM: **Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons**. *J Anim Sci* 2004, **82**(13).

21.     Abdurakhmonov IY, Abdukarimov A: **Application of association mapping to understanding the genetic diversity of plant germplasm resources**. *Int J Plant Genomics* 2008, **2008**.

22.     Berri C, Wacrenier N, Millet N, Le Bihan-Duval E: **Effect of selection for improved body composition on muscle and meat characteristics of broilers from experimental and commercial lines**. *Poult Sci* 2001, **80**:833-838.

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

24.     Montaldo HH, Meza-Herrera CA: **Use of molecular markers and major genes in the genetic improvement of livestock**. *EJB* 1998, **1**(2).

25.     Dekkers JCM, Hospital F: **The use of molecular genetics in the improvement of agricultural populations**. *Nature Reviews Genetics* 2002, **3**:22-32.

26.     Li H, Deeb N, Zhou H, Mitchell AD, Ashwell CM, Lamont SJ: **Chicken quantitative trait loci for growth and body composition associated with transforming growth factor-β genes1**. *Poult Sci* 2003, **82**:347-356.

27.     Nestor KE, McCartney MG, Harvey WR: **Genetics of growth and reproduction in the turkey. 1. Genetic and non-genetic variation in body weight and body measurements**. *Poult Sci* 1967, **46**:1374-1384.

28.     Toelle VD, Havenstein GB, Nestor KE, Bacon WL: **Estimates of genetic parameters in turkeys. 3. Sexual dimorphism and its implication in selection procedures**. *Poult Sci* 1990, **69**:1634–1643.

29.    Liu G, Dunnington EA, Siegel PB: **Maternal effects and heterosis for growth in reciprocal cross population of chickens**. *J Anim Breed Genet* 1993, **110**:423-428.

30.    Nestor KE, Anderson JW, Patterson RA: **Genetics of growth and reproduction in the turkey. 14. Changes in genetic parameters over thirty generations of selection for increased body weight**. *Poult Sci* 2000, **79**(4):445-452.

31.    Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: **A genome-wide scan for signatures of recent selection in Holstein cattle**. *Animal Genetics* 2010, **41**:377-389.

32.    Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al*: **Whole-genome resequencing reveals loci under selection during chicken domestication**. 2010, **464**(7288):587-591.

33.    Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC *et al*: **Genome sequence, comparative analysis and haplotype structure of the domestic dog**. 2005, **438**(7069):803-819.

34.    Pontius JU, Mullikin JC, Smith DR, Agencourt Sequencing Team, Lindblad-Toh K, Gnerre S, Clamp M, Chang J, Stephens R, Neelam B: **Initial sequence and comparative analysis of the cat genome**. *Genome Res* 2007, **17**:1675-1689.

35.    Nielsen R, Bustamante C: **A scan for positively selected genes in the genomes of humans and chimpanzees**. *Plos Biology* 2005, **3**(6):e170.

36.    Heil J, Glanowski S, Scott J, Winn-Deen E, McMullen I, Wu L, Gire C, Sprague A: **An automated computer system to support ultra high throughput SNP genotyping**. In: *Pacific symposium on biocomputing: 2002*; 2002: 7:30-40.

37.    Syvanen AC: **Accessing genetic variation: Genotyping single nucleotide polymorphisms**. *Nat Rev Genet* 2001, **2**:930-942.

38.    Derjusheva S, Kurganova A, Habermann F, Gaginskaya E: **High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds**. *Chromosome Res* 2004, **12**:715-723.

39.    Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W *et al*: **Whole-genome sequencing and variant discovery in *C. elegans***. *Nature Methods* 2008, **DOI:10:1038**.

40. Kerstens HHD, Crooijmans RPMA, Veenendaal A, Dibbits BW, Chin-A-Woeng TFC, Dunnen JTd, Groenen MAM: **Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey**. *BMC Genomics* 2009, **10**:479.

41. International Chicken Polymorphism Map Consortium: **A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms.** *Nature* 2004, **432**:717-722.

42. Brumfield RT, Beerli P, Nickerson DA, Edwards SV: **The utility of single nucleotide polymorphisms in inferences of population history**. *Trends in Ecol Evol* 2003, **18**:249-256.

43. Collins A, Lau W, De la Vega FM: **Mapping genes for common diseases: The case for genetic (LD) maps**. *Hum Heredity* 2004, **58**:2-9.

44. Bumstead N, Palyga J: **A preliminary linkage map of the chicken genome**. *Genomics* 1992, **13**(3):690-697.

45. Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden S, Hawkins GA, Toldo SS, Fries R, Grosz MD, Yoo J *et al*: **A genetic linkage map for cattle**. *Genetics* 1994, **136**:619-639.

46. Groenen MAM, Crooijmans RPMA, Veenendaal A, Cheng HH, Siwek M, Van der Poel JJ: **A comprehensive microsatellite linkage map of the chicken genome**. *Genomics* 1998, **49**:265-274.

47. Groenen MAM, Cheng HH, Bumstead N, Benkel BF, Briles WE, Burke T, Burt DW, Crittenden LB, Dodgson J, Hillel J *et al*: **A consensus linkage map of the chicken genome**. *Genome Res* 2000, **10**:137 - 147.

48. Burt DW, Morrice DR, Sewalem A, Smith J, Paton IR, Smith EJ, Bentley J, Hocking PM: **Preliminary linkage map of the turkey (*Meleagris gallopavo*) based on microsatellite markers**. *Anim Genet* 2003, **34**:399-409.

49. Jorgenson E, Tang H, Gadde M, Province M, Leppert M, Kardia S, Schork N, Cooper R, Rao DC, Boerwinkle E *et al*: **Ethnicity and human genetic linkage maps**. *Am J Hum Genet* 2005, **76**(2):276-290.

50. Reed KM, Sullivan LR, Foster LK, Chaves LD, Ponce de León FA: **Assignment of linkage groups to turkey chromosome 1 (MGA1)**. *Cytogenet Genome Res* 2006a, **115**:176-178.

51. Reed KM, Chaves LD, Knutson TP, Krueth SB, Ashwell CM, Burt DW: **Integration of microsatellite-based genetic maps for the turkey (*Meleagris gallopavo*)**. *Genome* 2006b, **49**:1308-1318.

52. Stapley J, Birkhead TR, Burke T, Slate J: **A linkage map of the Zebra Finch Taeniopygia Guttata provides new insights into avian genome evolution**. *Genetics* 2008, **179**:651-667.

53. Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens H-J, Crooijmans RPMA, Lathrop FBM, Muir WM, Ka-Shu Wong G, Gut I *et al*: **A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate**. *Genome Res* 2009, **19**(3):510-519.

54. Moore G, Devos KM, Wang Z, Gale MD: **Cereal genome evolution: Grasses, line up and form a circle**. *Current Biology* 1995, **5**(7):737-739.

55. Doganlar S, Frary A, Daunay M-C, Lester RN, Tanksley SD: **A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae**. *Genetics* 2002, **161**:1697-1711.

56. Moore G, Devos KM, Wang Z, Gale MD: **Cereal Genome Evolution - Grasses, Line up and Form a Circle**. *Curr Biol* 1995, **5**:737- 739.

57. Elferink M, van As P, Veenendaal T, Crooijmans R, Groenen M: **Regional differences in recombination hotspots between two chicken populations**. *BMC Genetics* 2010, **11**(1):11.

58. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ: **Comparative recombination rates in the rat, mouse, and human genomes**. *Genome Res* 2004, **14**(4):528-538.

59. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MAM, Delany ME *et al*: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution**. *Nature* 2004, **432**:695 - 716.

60. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G *et al*: **A high-resolution recombination map of the human genome**. *Nat Genet* 2002, **31**(3):241-247.

61. Reed KM, Chaves LD, Hall MK, Knutson TP, Harry DE: **A comparative genetic map of the turkey genome**. *Cytogenet Genome Res* 2005, **111**:118-127.

62. Griffin D, Robertson L, Tempest H, Vignal A, Fillon V, Crooijmans R, Groenen M, Deryusheva S, Gaginskaya E, Carre W *et al*: **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution**. *BMC Genomics* 2008, **9**(1):168.

63. Griffin DK, Robertson LBW, Tempest HG, Skinner BM: **The evolution of the avian genome as revealed by molecular cytogenetics**. *Cytogenet Genome Res* 2007, **117**:64-77.

64. Gelderman H: **Investigation on inheritance of quantitative characters in animals by gene markers**. *I Methods Theoretical and Applied Genetics* 1975, **46**:300-319.

65. Meuwissen THE, Goddard ME: **The use of marker-haplotypes in animal breeding schemes**. *Genetics Selection Evolution* 1996, **28**:161-176.

66. Pyasatian N, Fernando RL, Dekkers JCM: **Genomic selection for marker-assisted improvement in line crosses**. *Theor Appl Genet* 2007, **115**:665-674.

67. van Kaam JBCHM, Groenen MAM, Bovenhuis H, Veenendaal A, Vereijken ALJ, van Arendonk JAM: **Whole genome scan in chickens for quantitative trait loci affecting growth and feed efficiency**. *Poult Sci* 1999a, **78**:15-23.

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

69. Nadaf J, Gilbert H, Pitel F, Berri CM, Feve K, Beaumont C, Duclos MJ, Vignal A, Porter TE, Simon J *et al*: **Identification of QTL controlling meat quality traits in an F2 cross between two chicken lines selected for either low or high growth rate**. *BMC Genomics* 2007, **8**:155.

70. Wahlberg P, Carlborg Ö, Foglio M, Tordoir X, Syvänen A-C, Lathrop M, Gut IG, Siegel PB, Andersson L: **Genetic analysis of an F2 intercross between two chicken lines divergently selected for body-weight**. *BMC Genomics* 2009, **10**:248.

71. Ankra-Badu GA, Le Bihan-Duval E, Mignon-Grasteau S, Pitel F, Beaumont C, Duclos MJ, Simon J, Carre W, Porter TE, Vignal A *et al*: **Mapping QTL for growth and shank traits in chickens divergently selected for high or low body weight**. *Ani Genet* 2010, **41**:400-405.

72. Yonash N, Bacon LD, Witter RL, Cheng HH: **High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease**. *Anim Genet* 1999, **30**:126-135.

73. van Kaam JBCMH, Groenen MAM, Bovenhuis H, Veenendaal A, Vereijken ALJ, van Arendonk JAM: **Whole genome scan in chickens for quantitative trait loci affecting carcass traits**. *Poult Sci* 1999b, **78**:1091-1099.

74. Yonash N, Cheng HH, Hillel J, Heller DE, Cahaner A: **DNA microsatellites linked to quantitative trait loci affecting antibody response and survival rate in meat-type chickens**. *Poult Sci* 2001, **80**:22-28.

75. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain terminating inhibitors.** *Proc Natl Acad Sci U S A* 1997, **74**:5463-5467.

76. Mardis ER: **Next-Generation DNA sequencing methods.** *Annual Review of Genomics and Human Genetics* 2008, **9**:387-402.

77. Pushkarev D, Neff NF, Quake SR: **Single-molecule sequencing of an individual human genome.** *Nat Biotech* 2009, **27**:847-850.

78. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES: **An SNP map of the human genome generated by reduced representation shotgun sequencing**. *Nature* 2000, **407**(6803):513-516.

79. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al*: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nature Genetics* 2011, **43**(5):491-498.

80. Li H, Homer N: **A survey of sequence alignment algorithms for next-generation sequencing**. *Briefings in Bioinformatics* 2010, **11**(5):473-483.

81. Stromberg M: **Mosaik Assembler**. In: *110014.* Edited by Lee W-P: Boston College; 2010.

82. Gusella JF, Tanzi RE, Bader PI, Phelan MC, Stevenson R, Hayden MR, Hofman KJ, Faryniarz AG, Gibbons K: **Deletion of Huntington's disease-linked G8 (D4S10) locus in Wolf-Hirschhorn syndrome**. *Nature* 1985, **318**(6041):75-80.

83. D'Haene B, Attanasio C, Beysen D, Dostie J, Lemire E, Bouchard P, Field M, Jones K, Lorenz B, Menten B *et al*: **Disease-Causing 7.4 kb *Cis*-Regulatory Deletion Disrupting Conserved Non-Coding Sequences and Their Interaction with the *FOXL2* Promotor: Implications for Mutation Screening**. *PLoS Genet* 2009, **5**(6):e1000522.

84. Ellegren H: **Molecular evolutionary genomics of birds**. *Cytogenetic and Genome Research* 2007, **117**(1-4):120-130.

85. Megens HJ, Crooijmans RP, San Cristobal M, Hui X, Li N, Groenen MA: **Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication**. *Genet Sel Evol* 2008, **40**(1):103-128.

86. Ellegren H: **Evolutionary stasis: the stable chromosomes of birds**. *Trends in Ecology & Evolution* 2010, **25**(5):283-291.

87. Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MAM: **Genome-Wide Footprints of Pig Domestication and Selection Revealed through Massive Parallel Sequencing of Pooled DNA**. *PLoS One* 2011, **6**(4):e14782.

88. Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, Thaller G, Simianer H: **Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle**. *BMC Genomics* 2011, **12**:318.

89. Chun S, Fay JC: **Identification of deleterious mutations within three human genomes**. *Genome Res* 2009, **19**:1553-1561.

90. Davidson S, Starkey A, MacKenzie A: **Evidence of uneven selective pressure on different subsets of the conserved human genome; implications for the significance of intronic and intergenic DNA**. *BMC Genomics* 2009, **10**:614.

91. Neff BD: **Stabilizing selection on genomic divergence in a wild fish population**. *PNAS* 2004, **101**(8):2381-2385

92. Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A: **Patterns of Positive Selection in Six Mammalian Genomes**. *PLoS Genet* 2008, **4**(8):e1000144.

93. Montoya-Burgos JI: **Patterns of Positive Selection and Neutral Evolution in the Protein-Coding Genes of *Tetraodon* and *Takifugu***. *PLoS ONE* 2011, **6**(9):e24800.

94. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Ann Blomberg L, Bouffard P, Burt DW, Crasta O, Crooijmans RPMA *et al*: **Multi-Platform Next-Generation Sequencing of the Domestic Turkey (*Meleagris gallopavo*): Genome Assembly and Analysis**. *PLoS Biol* 2010, **8**(9):e1000475.

# 2

# Genetic variances, heritabilities and maternal effects on body weight, breast meat yield, meat quality traits and the shape of the growth curve in turkey birds

Muhammad L Aslam[1], John WM Bastiaansen[1], Richard PMA Crooijmans[1], Bart J Ducro[1], Addie Vereijken[2], Martien AM Groenen[1]

[1] Animal Breeding and Genomics Centre, Wageningen University, 6709PG, Wageningen, the Netherlands

[2] Hendrix Genetics, Research & Technology Centre, 5830 AC, Boxmeer, the Netherlands

## Abstract

### Background

Turkey is an important agricultural species and is largely used as a meat bird. In 2004, turkey represented 6.5% of the world poultry meat production. The world-wide turkey population has rapidly grown due to increased commercial farming. Due to the high demand for turkey meat from both consumers and industry global turkey stocks increased from 100 million in 1970 to over 276 million in 2004. This rapidly increasing importance of turkeys was a reason to design this study for the estimation of genetic parameters that control body weight, body composition, meat quality traits and parameters that shape the growth curve in turkey birds.

### Results

The average heritability estimate for body weight traits was 0.38, except for early weights that were strongly affected by maternal effects. This study showed that body weight traits, upper asymptote (a growth curve trait), percent breast meat and redness of meat had high heritability whereas heritabilities of breast length, breast width, percent drip loss, ultimate pH, lightness and yellowness of meat were medium to low. We found high positive genetic and phenotypic correlations between body weight, upper asymptote, most breast meat yield traits and percent drip loss but percent drip loss was found strongly negatively correlated with ultimate pH. Percent breast meat, however, showed genetic correlations close to zero with body weight traits and upper asymptote.

### Conclusion

The results of this analysis and the growth curve from the studied population of turkey birds suggest that the turkey birds could be selected for breeding between 60 and 80 days of age in order to improve overall production and the production of desirable cuts of meat. The continuous selection of birds within this age range could promote high growth rates but specific attention to meat quality would be needed to avoid a negative impact on the quality of meat.

## 2.1 Background

Turkey is largely used as a meat bird. In 2004, turkey represented 6.5% of the world poultry meat production [1]. The world-wide turkey population has rapidly grown due to increased commercial farming. Global turkey stocks increased from 100 million in 1970 to over 276 million in 2004. Over the same time period, the production volume increased from 1.2 to 5.1 million [1]. Due to the high demand for turkey meat from both consumers and industry, the breeding objective is to produce rapidly growing birds with a high market body weight (BW) and a desirable body conformation in order to maximize production efficiency and optimize production of preferred body cuts; e.g., breast muscle [2]. These objectives can be achieved by selective breeding of birds for high body weights, with much emphasis on breast muscle yield, while considering the efficiency of production over the growth curve. Knowledge of the growth curve will be useful when defining ages and weights at which to select birds as well as for the design of management procedures.

Breeding programs for meat type birds are commonly selecting for BW, and body composition traits (breast yield, etc.) while minimizing production costs. Recently breeders have started to measure meat quality (drip loss, pH, etc.) as well as survival traits, at least in research projects [3, 4]. Selection was found successful to improve growth and body composition traits while these traits did not show any negative association with the excessive drip loss in chicken [3, 5]. Drip loss was found correlated with pH of meat and differences in pH significantly affect the storage and the processing quality of the meat [5, 6]. Meat with low pH is characterized by a low water-holding capacity and poor technological quality and is therefore referred to as pale, soft, and exudative (PSE) meat [6, 7]. Meat with high pH, known as dry, firm, and dark (DFD) meat, is characterized by a poor storage quality which is the result of a faster rate of off-odor production and an accelerated microbiological growth [8].

BW traits were found to be influenced by not only genetics but also common or maternal environmental effects [9]. Nestor et al. [10] reported that the un-weighted averages of published narrow sense heritability ($h^2$) estimates of BW in selected populations of turkey birds were 0.40, 0.42, 0.43, and 0.36 for birds in the age groups 0 to 8, 9 to 16, 17 to 24, and over 24 wk, respectively [10]. Other studies also found high heritabilities for BW at various ages, ranging from 0.28 to 0.48 [11-13]. Strong positive genetic correlations were found between the 16-wk BW and BW at other ages (8, 20, and 24 wk of age). Negative correlations were found between BW and reproduction traits [13]. Toelle et al. [14] found that the genetic correlation between BW of the two sexes at 16 wk of age was close to unity.

Many reports exist that show estimation of growth curves; an understanding of growth curves is important for the efficient production of animals [15]. Growth curve parameters were estimated for turkeys by Sengul and Kuraz [16] with four different non-linear models (Gompertz, Logistic, Morgan-Mercer-Flodin [MMF], and Richards) and very good fits were found with Gompertz, Logistic and Richards models. Mignon-Grasteau et al. [17] estimated growth curve parameters with the Gompertz function in chickens. High heritabilities were found for these growth curve parameters [17]. It was established that the growth curve varies among individuals; thus, growth might be enhanced by selection on the basis of growth curve parameters [18].

In this study, we estimated genetic parameters for different growth (BW and growth curve traits) and meat quality traits in turkeys as well as the genetic and phenotypic correlations between these traits. To our knowledge, this is the first study to evaluate genetic parameters for meat quality in turkey and to estimate correlations of turkey meat quality with growth traits and meat yield traits.

## 2.2 Methods

### *Animals*

The study population was based on two genetically different commercial turkey lines referred to as line A and line B. Line A was selected for rapid growth and line B was selected for a high reproduction rate. Males from line A were crossed to females from line B to produce F1 offspring. From the F1 generation, 25 males and 34 females were randomly selected and mated to produce 1,716 F2 offspring. The number of F2 offspring in a full-sib group ranged from 16 to 120 with an average of 63 offspring per group. Each F1 female was mated once; therefore the pedigree included no maternal half-sibs. F2 individuals had pedigree information for 9 generations and phenotypes were recorded only on F2 individuals. The pedigree consisted of 2,186 individuals; the F2 individuals were from 14 different hatch dates between 21-05-2000 and 04-11-2001. The package pedigree, in R statistical software [19], was used to check the pedigree file for potential errors.

### *Feeding Schedule*

Turkey birds were fed according to the feed schedule and nutrient guidelines of Hybrid (A Hendrix Genetics Company). Feed changed in energy (ME/Kg), crude protein percent (CP) and other essential nutrients level with the age of a bird. Energy level of feed was raised while CP level was lowered with increasing age of birds. In the 1st week, feed was supplied with a CP level of 27.5 % and an energy level of 2850 ME/Kg while in the 17th week of age CP level had been lowered to 17 % and energy level had been raised to 3520 ME/Kg.

### *Housing Conditions*

Turkey birds were raised in unisex groups of around 500 to 525 poults/group. The bedding material was comprised of wood shavings for the entire rearing period, and in the first week of age, brooder rings were used. The birds were kept in closed barns with concrete floors and controlled lighting and ventilation systems. The same duration of light (12 hr/day) was provided to both male and female birds during the first 15 weeks. After 15 weeks, light was provided for 14 hr/day and 16 hr/day to male and female birds, respectively. The environmental temperature was maintained at a relatively high level of 22.8 to 27.8°C during the first week, after which it was decreased gradually with the age of the birds. After 12 weeks, the temperature was kept constant at 13.9 to 16.1°C. In the first 6 weeks, birds were provided floor space of 0.074m$^2$/bird. After 6 weeks, the floor space was increased to 0.167 m$^2$/female and 0.185 m$^2$/male up to 15 weeks; the final floor space of 0.209 m$^2$/female and 0.269 m$^2$/male was provided during 16 to 20 weeks of age.

### *Traits*

Phenotypic data were recorded as part of a commercial breeding program. BW and carcass related traits were recorded for 1,716 (692 females and 1,024 males) individuals of the F2 generation. Body weights were recorded at 1, 17, 40, 60, 80, and 120 days (BW01, BW17, BW40, BW60, BW80, and BW120, respectively). The breast meat yield traits breast length (BrL) and breast width (BrW) were measured with a caliper in live birds just before slaughter at 20 weeks of age. BrW was measured at the widest point of the breast while BrL was measured at the symmetry line of the breast. The percent breast meat (PBM) and percent drip loss (PDL) were recorded at 20 weeks of age after the birds were slaughtered. PDL was measured in breast meat samples of 30 to 50g. After measuring initial weight, samples were packed and hung for five days at a temperature of 4°C. After a storage period of five days, the samples were weighed again for the final weights. The PDL was recorded as a percentage of initial weight [20].

The ultimate pH (pHu) of the *Pectoralis major* muscle of a skinless breast fillet was measured at 24 h post-slaughter with a piercing electrode (Cole Parmer L-05992-22, Chicago, Illinois). Breast meat color was measured at 24 h post-slaughter using a portable Minolta Chroma Meter (Model CR-200; Ramsey, NJ) with the CIE L*a*b* system, where L* represents lightness, a* redness and b* yellowness. Higher L*, a* and b* values correspond to paler, redder and more yellow meat, respectively. The Minolta Chroma Meter was calibrated according to the CIELAB color system. The pH and color were measured in the same area of the breast, on the thickest position of the lobe.

*Growth Curve*

Growth curve parameters were estimated with a logistic growth function (SSlogis) in R statistical software [21]. Only individuals that had measurements for BW01 and BW120 and at least 2 additional BW measurements were included for the estimation of growth curve parameters. With these restrictions 867 out of the total 1,716 birds were included. Population parameter values of the logistic growth curve were estimated for the male and female populations separately as well as sex average parameter values. Growth curves were plotted for every individual in a population using their estimated parameter values. Separate logistic growth curves were also plotted for the male and female populations as well as the complete population with the estimated parameter values. To estimate the parameters of the logistic growth curve, the following equation was fitted to the data:

$$W_t = \frac{As_{wt}}{1 + e^{t_{mid} - t / scale}}$$

where W(t) is weight at time t (days), $As_{wt}$ is the asymptotic weight (Kg), $t_{mid}$ is the inflection point at which 50% of the asymptotic weight is achieved (days), and **scale** is a constant that is proportional to the overall growth rate [22, 23].

*Genetic Analyses*

Descriptive statistics were obtained from a generalized linear model (PROC GLM [24]). The correction of data and removal of outlier values (>3 SD) and the test for the normality of the distribution of traits was performed with method PROC UNIVARIATE [24]. Only PBM and PDL displayed outlier values (>3 SD) and those animals were removed from the analysis. Fixed effects of sex and hatch date were tested for significance of their effect on each trait with PROC GLM [24]. Effects that were found significant ($P < 0.05$) were included in the model for the estimation of genetic parameters.

Heritabilities for all the traits under study were estimated with an animal model in ASREML statistical software [25] using univariate analyses. Bivariate analyses for all possible combinations of traits were applied to estimate genetic and phenotypic correlations. Estimates obtained in univariate analysis were used as starting values in bivariate analyses. In the ASREML program, the maximum number of iterations was set to 20; for the most part, convergence criteria were met in less than 10 iterations and always before 20 iterations. An additional 10 iterations after convergence did not change results. Convergence was presumed when the log-likelihood changed less than 0.002 between iterations and the individual variance parameter estimate changed less than 1% [25].

In addition to the genetic analyses mentioned above, the genetic correlations between BW of males and BW of females at the same age (e.g. BW01M and BW01F) were also estimated for each BW trait using a bivariate analysis to test if male and female growth should be regarded separate traits.

A random common environment effect of the dam was included in the model for all the traits, except for meat quality traits (PDL, pHu, L*, a* and b*). A likelihood ratio test (**LR-test**) was used to check the significance of the full model (with a random common environment of dam) compared to the reduced model (without a random common environment of dam) based on the following equation:

$$Y_{ijkl} = \mu + S_i + H_j + A_k + C_l + E_{ijkl}$$

Where $Y_{ijkl}$ is the performance of individual $k$, $\mu$ is overall mean, $S_i$ is the fixed effect of sex $i$, $H_j$ is the fixed effect of the week of hatch $j$ ($j$= 1, 2...14), $A_k$ is the random direct genetic effect of individual $k$ with $a \sim N(0, A\sigma_a^2)$, $C_l$ is the random common environment effect of the $l$-th dam with $c \sim N(0, I\sigma_c^2)$, and $E_{ijkl}$ is the random residual effect.

### *Ethical approval for the use of animals in this study*

Although animals were used in this work, no experiments were performed on them. Data was recorded as a part of the routine work at a breeding company (Hendrix Genetics). No approval from the ethics committee was necessary.

## 2.3 Results

### *Descriptive Analysis*

A descriptive analysis of all the traits studied was summarized in Table 2.1. The effect of sex was significant ($P < 0.05$) for all the traits except for the weight of 1 day old chicks (BW01) and the redness of meat (a*). The mean values for all the traits studied were higher for males than females. The effect of hatch date was also significant for all the traits.

**Table 2.1:** Descriptive statistics, including the estimates for the significant fixed effects (Sex and Hatch).

| Traits(units) | N | Minimum | Maximum | LS Mean | RSD | Sex[1] | Hatch[2] |
|---|---|---|---|---|---|---|---|
| BW01(Kg) | 1416 | 0.04 | 0.07 | 0.06 | 0.00 | 0.00 | 0.02*** |
| BW17(Kg) | 1281 | 0.08 | 0.60 | 0.35 | 0.05 | 0.21*** | 0.11*** |
| BW40(Kg) | 1226 | 0.52 | 2.32 | 1.40 | 0.17 | 0.66*** | 0.37*** |
| BW60(Kg) | 1103 | 1.50 | 4.96 | 3.18 | 0.37 | 1.27*** | 0.65*** |
| BW80(Kg) | 1009 | 3.04 | 8.50 | 5.57 | 0.59 | 2.22*** | 1.64*** |
| BW120(Kg) | 878 | 4.54 | 15.90 | 10.49 | 1.01 | 5.00*** | 1.48*** |
| PBM (%) | 919 | 9.10 | 13.40 | 11.19 | 0.71 | -0.20*** | 1.17*** |
| BrL(mm) | 1198 | 149.00 | 249.00 | 196.35 | 14.08 | 49.44*** | 23.19*** |
| BrW(mm) | 1198 | 107.00 | 181.50 | 135.93 | 7.91 | 26.70*** | 21.80*** |
| PDL(%) | 1028 | 2.21 | 14.10 | 5.11 | 1.14 | 0.94*** | 1.36*** |
| pHu | 1055 | 5.22 | 6.08 | 5.73 | 0.09 | 0.03*** | 0.49*** |
| L* | 1083 | 40.30 | 53.60 | 45.94 | 1.72 | 1.03*** | 2.65*** |
| a* | 1083 | 1.30 | 9.20 | 5.25 | 0.97 | 0.06 | 1.06*** |
| b* | 1083 | 0.00 | 5.60 | 2.25 | 0.77 | 0.51*** | 0.81*** |
| $As_{wt}$ (Kg) | 867 | 4.6 | 20.23 | 12.39 | 1.32 | 6.47*** | 2.82*** |
| $t_{mid}$(day) | 867 | 59.86 | 112.24 | 82.82 | 3.58 | 6.07*** | 13.20*** |
| scale(day) | 867 | 12.66 | 29.15 | 20.61 | 1.21 | 1.86*** | 5.78*** |

N = Number of records; minimum = minimum values; maximum = maximum values; LS Mean = least square mean; RSD = residual standard deviation; BW01, BW17, BW40, BW60, BW80 ,and BW120 are the BW at days 1,17, 40, 60, 80, and 120 of age, respectively; PBM = percentage breast meat at 20 week of age; BrL = breast length at 20 week of age; BrW = breast width at 20 wk of age; PDL = percent drip loss at 20 week of age; pHu = ultimate pH at 20 wk of age; L* = lightness at 20 wk of age; a* = redness at 20 wk of age; b* = yellowness at 20 wk of age; $As_{wt}$ = upper asymptotic weight (estimated growth curve parameter); $t_{mid}$ = inflection point at 50% asymptote (estimated growth curve parameter); scale = constant that is proportional to the overall growth rate (estimated growth curve parameter). [1] = The difference between sexes in the Least square means (LS Means) of the traits. [2] = The difference between the maximum and minimum LS Means of the traits with respect to the week of hatch. *$P \le 0.05$, **$P \le 0.005$, ***$P \le 0.0005$.
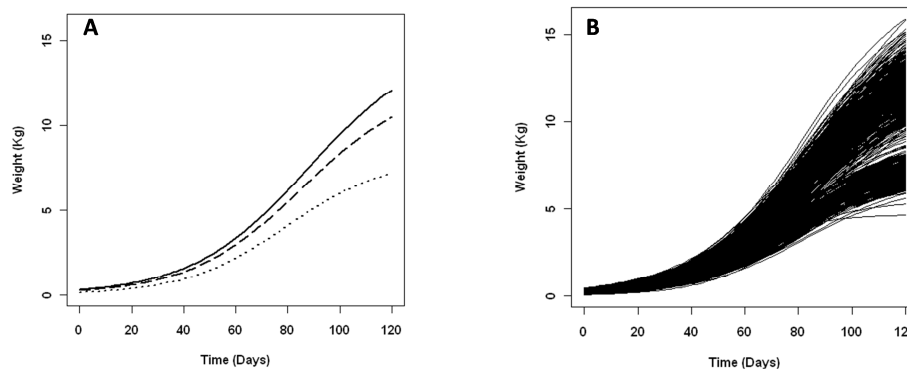
### Growth Curve

The average parameter values estimated from the logistic growth curve are given in Table 2.2. The logistic growth curves were estimated and plotted from actual measurements of BW throughout the growth period; in this case, BW01, BW17, BW40, BW60, BW80, and BW120 (Figure1). The male and female populations showed a difference in growth rate that was apparent in the estimates of the growth curve parameters and could also be observed in Figure 2.1B which shows an apparent split into 2 groups of the individual animal growth curves.

**Table 2.2:** Estimates of logistic growth curve parameters for males, females, and sex average parameter values.

|  | $As_{wt}$(Kg) | $t_{mid}$(day) | scale(day) |
|---|---|---|---|
| **Male** | 14.44 | 84.87 | 21.39 |
| **Female** | 7.88 | 78.28 | 19.22 |
| **Sex average** | 11.16 | 81.58 | 20.31 |

$As_{wt}$ = upper asymptote (estimated growth curve parameter); $t_{mid}$ = inflection point at 50% asymptote (estimated growth curve parameter); scale = constant that is proportional to the overall growth rate (estimated growth curve parameter).

**Figure 2.1:** Logistic growth curves depicting the change in growth rate of the turkey population through time.



**A** = Growth curves representing average growth rates in males (solid line), females (dotted line), and the whole population (dashed line); **B** = Growth curves of all the individuals in the population.

### *Heritability Estimates*

Body weight traits BW40, BW60, BW80, and BW120 were found to be highly heritable, with heritability estimates ($h^2$) of 0.32, 0.39, 0.42, and 0.40, respectively (Table 2.3). The BW at 1 and 17 days (BW01 and BW17) were found to have low heritability, with estimates of 0.0 and 0.12 respectively. The proportion of variance explained by common (maternal) environment was 0.43 at BW01. This proportion reduced rapidly to 0.11 at BW17 and became negligible after BW60.

The heritability estimates for breast meat yield traits PBM, BrL, and BrW were in the moderate to high range, with estimates of 0.30, 0.15, and 0.17, respectively. The meat quality traits PDL, pHu, L*, a* and b* showed low to high estimates of heritability. PDL and pHu showed low heritabilities of 0.12 and 0.09 respectively while the other quality traits L*, a* and b* showed moderate to high heritabilities with estimates at 0.27, 0.30 and 0.15 respectively.

41

The growth curve trait $As_{wt}$ showed a high heritability estimate of 0.30, and the remaining two growth curve traits, $t_{mid}$ and scale, showed lower heritabilities at 0.05 and 0.11, respectively.

**Table 2.3:** Estimates of heritability, standard deviations and common environment variance ratios for different traits.

| Trait | $\sigma_a$ | $\sigma_c$ | $h^2$ (S.E) | $c^2$ (S.E) |
|---|---|---|---|---|
| BW01(Kg)[1] | 0.0045 | 2.99 | 0.00(0.00) | 0.43(0.06) |
| BW17(Kg)[1] | 18.25 | 17.11 | 0.12(0.20) | 0.11(0.08) |
| BW40(Kg) | 99.60 | 36.36 | 0.32(0.22) | 0.04(0.07) |
| BW60(Kg) | 241.17 | 41.13 | 0.39(0.26) | 0.01(0.08) |
| BW80(Kg) | 396.29 | 0.06 | 0.42(0.12) | 1.09E-08(0.00) |
| BW120(Kg) | 652.68 | 0.04 | 0.40(0.12) | 1.86E-09(0.00) |
| PBM(%) | 0.39 | 7.11E-10 | 0.30(0.10) | 1.36E-09(0.00) |
| BrL(mm) | 5.51 | 4.80E-06 | 0.15(0.06) | 2.37E-08(0.00) |
| BrW(mm) | 3.30 | 1.12E-07 | 0.17(0.07) | 1.73E-09(0.00) |
| PDL(%) | 0.40 | NI | 0.12(0.06) | NI |
| pHu | 0.03 | NI | 0.09(0.04) | NI |
| L* | 0.90 | NI | 0.27(0.09) | NI |
| a* | 0.54 | NI | 0.30(0.09) | NI |
| b* | 0.30 | NI | 0.15(0.05) | NI |
| $As_{wt}$(Kg) | 737.68 | 0.07 | 0.30(0.10) | 2.67E-09(0.00) |
| $t_{mid}$(day) | 0.80 | 0.00021 | 0.05(0.04) | 3.44E-09(0.00) |
| scale(day) | 0.41 | 0.00037 | 0.11(0.06) | 8.99E-08(0.00) |

$\sigma_a$ = Additive genetic standard deviation; $\sigma_c$ = common environment standard deviation; $h^2$ (S.E) = heritability estimates with standard errors (S.E); $c^2$ (S.E) = common environment variance ratio with standard errors (S.E); BW01, BW17, BW40, BW60, BW80, and BW120 are the BW at days 1,17, 40, 60, 80, and 120 of age; PBM = percentage breast meat at 20 wk of age; BrL = breast length at 20 wk of age; BrW = breast width at 20 wk of age; PDL = percent drip loss at 20 wk of age; pHu = ultimate pH at 20 wk of age; L* = lightness at 20 wk of age; a* = redness at 20 wk of age; b* = yellowness at 20 wk of age; $As_{wt}$ = upper asymptote (estimated growth curve parameter); $t_{mid}$ = inflection point at 50% asymptote (estimated growth curve parameter); scale = constant that is proportional to the overall growth rate (estimated growth curve parameter). NI = Not Included (common environment) in the analysis for the trait. [1] = Full model with common environment effect was found significantly different from the reduced model for these traits; $P < 0.05$.

### *Genetic and Phenotypic Correlations*

Genetic and phenotypic correlations were estimated between all the BW traits (BW17, BW40, BW60, BW80, and BW120), except BW01, which showed zero heritability. We found high positive genetic correlations among all the BW traits ranging from 0.86 to 0.98 (Additional file 2.1). Genetic correlations decreased as

the time between BW measurements increased, except for correlations with BW120. At this point, the birds were well past the inflection point (Figure 2.1) and were close to their final adult BW. Phenotypic correlations among all the BW traits were also found to be high and positive. Genetic correlations between BW of males and BW of females at the same age were found to be high in the range of 0.87 – 0.99 for all BW traits. BW measures were therefore treated as one trait in subsequent analyses.

Positive genetic correlations were also found among the breast meat yield traits, BrL, BrW, and $As_{wt}$ which ranged from 0.68 to 0.90. These traits also showed high positive phenotypic correlations. All BW traits and $As_{wt}$ showed genetic correlations close to zero with PBM, albeit with large standard error of estimates. Positive phenotypic correlations of PBM with BW traits and $As_{wt}$ ranged from 0.21 to 0.32. The $As_{wt,}$ BrL, and BW traits showed positive genetic and phenotypic correlations with PDL, lightness and yellowness (L* and b*). The traits pHu and a* showed negative genetic correlation with $As_{wt,}$ BrL, and BW traits but these results had a high standard error of estimates. Phenotypic correlation of pHu and a* with $As_{wt,}$ BrL, and BW traits was close to zero. The ultimate pH had negative genetic and phenotypic correlations with PDL with the genetic correlation estimated close to minus one (Additional file 2.1). PDL showed positive genetic and phenotypic correlations with L*, a* and b* while L* had negative genetic and phenotypic correlations with a* and positive genetic and phenotypic correlations with b*. The genetic and phenotypic correlations between a* and b* were close to zero. The growth curve parameter $t_{mid}$ showed a highly negative genetic correlation with PDL, and phenotypic correlations that were either negative or close to zero with all other traits except for the other two growth curve traits $As_{wt}$ and scale. Genetic and phenotypic correlations between the PBM and PDL were close to zero.

Positive genetic and phenotypic correlations were found among PBM, BrL, and BrW. BrL showed positive genetic and phenotypic correlations with PDL. In contrast, BrW showed a negative genetic correlation and a positive phenotypic correlation with PDL. All the correlations of BrL and BrW with PDL were however close to zero (Additional file 2.1).

The analysis of growth curve traits showed that $As_{wt}$ had negative genetic correlations, but positive phenotypic correlations, with $t_{mid}$ and scale. Positive genetic and phenotypic correlations were observed between $t_{mid}$ and scale.

## 2.4 Discussion

The aim of this study was to estimate heritabilities and determine genetic and phenotypic correlations for BW, breast meat yield, and meat quality traits in

turkeys. We also aimed to estimate the growth curve and the heritabilities of its parameters. The phenotypes used in this study were measured on an F2 cross between 2 turkey lines with a different genetic background and selected for different traits. The variances obtained are relevant to the F2 cross and cannot be directly applied to existing breeding stock. The estimates do provide a useful benchmark for breeders interested in the potential for correlated responses in meat quality from selection on growth and yield and for breeders who contemplate the estimation of heritabilities in their breeding lines and/or adding these traits to their breeding objectives.

In the present study, body weight was considered to be a single trait across both sexes, with sex used as a fixed effect in the analyses. This was in contrast with other studies, where parameters were estimated separately for males and females [13, 26, 27]. Parameters were not estimated separately in our analyses because those estimates would have been based on a subset of our relatively small population. Joint analysis of males and females seems warranted because genetic correlations between BW of males and BW of females at the same age were found to be high. In addition to sex, hatch date was included as a fixed effect in the analyses because it was found to play a significant role in BW and other traits in the study [28, 29].

In the present study, univariate models were used for the estimation of heritability and bivariate models for the estimation of genetic and phenotypic correlations [30]. Multivariate analyses were performed for small groups of related traits and results were not different from those obtained from univariate and bivariate models. Combining traits did not always result in convergence of the REML estimation. A common environmental variance ($c^2$) was found significant for some traits (BW01 & BW17) and not for others which further complicated the estimation from multivariate models.

We found heritability estimates for BW traits in the expected range, except for BW01 and BW17, which is attributed to the strong common environment effect at those early ages. Results are in range with previously reported heritability estimates. BW traits at various ages were reported to have an average heritability of 0.41 in a review of eighteen reports by Arthur and Abplanalp [11]. Similar results were also reported by Buss [31], who observed heritability in the range of 0.23 to 0.71 for BW traits at different ages.

The common environment effect had a large impact on the estimates of heritability for BW traits, especially at early ages. Neglecting the common environment effect would have resulted in an overestimation of heritabilities at early ages. For comparison, we estimated heritabilities without including the common

environmental effect (results not shown), and found that the estimated heritability of body weight was increased at all ages, but especially for BW01 and BW17. Similar conclusions were reached by others regarding the effect of common environment on the estimation of heritability [12, 32-34]. In our study, $c^2$ was found to decrease with increases in age and it was close to zero at later ages. The direct genetic component was found to increase with age which could be attributed to the initiation of expression of the animal's own genetics.

In the present study, the BW of day old turkey chicks had a heritability close to zero. Tullett and Burton [35] found in a study on broilers that 97% of the variation in chick weight at hatching was due to two factors: fresh egg weight and weight loss during incubation. Moreover, North [36] found that egg weight represented 70% of the chick weight. Taken together, these results suggest that day old BW was not heritable, but egg weight or egg size was heritable.

Our heritability estimates of the other production traits, including PBM, BrL, and BrW, were also consistent with reports from other groups. Our heritability estimate for PBM was 0.30, similar to values found by Le Bihan-Duval et al. [2] in chickens. The comparison is made to chicken because it is the closest related species to the turkey for which values are available. Our heritability estimates for breast length and breast width were low and quite close to each other. These results were in agreement with the work of Adeyinka et al. [28] on chickens. Our heritability estimate for PDL at 0.12 was the first reported for turkey meat, and somewhat inconsistent with the heritability of 0.26 found in chickens by Le Bihan-Duval et al. [2]. Besides the estimate being made in different species there were also differences in the measurement of traits with Le Bihan-Duval et al. [2] measuring PDL from the whole breast muscle while a smaller breast meat sample was used in our study.

The heritabilities in the present study for pHu, a* and b* at 0.09, 0.30 and 0.15 were found roughly in agreement with the results of Le Bihan-Duval et al. [37] in turkeys, while our estimate of heritability for L*, 0.27, was somewhat higher that the value of 0.12 obtained Le Bihan-Duval et al. [37]. A possible explanation can be sought in the different fixed effects included in the models by these two studies which in turkey may have explained a bigger part of the residual variance for this particular trait L* .

Sengul and Kuraz [16] concluded that Gompertz, Logistic, and Richards models all performed well for describing growth in turkeys. The logistic and Gompertz models have fixed growth forms with points of inflection at about 50 and 37% of the asymptote, respectively [22]. These parameter models are special cases of the more flexible Richards model, which has a variable point of inflection specified by

45

the shape parameter [38]. The growth models (Logistic, Gompertz and Richards) also differ slightly from each other in the interpretation of other parameters [39]. Here, we choose to use the logistic growth model for the analyses of growth. The $As_{wt}$ (upper asymptote) had high heritability, consistent with that found by Mignon-Grasteau et al. [17] who used the Gompertz model in chickens. We found low heritability estimates for $t_{mid}$ and scale which was not in agreement with the results reported by Grossman and Bohren [40] in chicken but the heritability estimate for $t_{mid}$ from our study was in agreement with the results from Le Rouzic et al. [41] in chicken who used a Gompertz growth model. Inconsistency in the results of the present study and the study by Grossman and Bohren [40] for $t_{mid}$ and scale could be due to the difference in species, differences between methods for the estimation of genetic parameters (based on correlation among full-sibs in Grossman and Bohren [40]) or because of the high margin of error reported in the study by Grossman and Bohren [40]. The differences we observed between the estimates of growth curve parameters for males and females were similar to differences observed by Sengul and Kuraz [16] in white turkeys and by Barbato and Younken [42] in chickens.

In the present study, the genetic correlations among all the BW traits ranged from 0.86 to 0.99. Genetic correlations were higher for measurements taken close together in age and declined somewhat as the measurement were taken farther apart in age. Similar results on genetic correlations among multiple BW traits were reported by Kranis et al. and Chapuis et al. [12, 26], who applied various mixed models and performed multivariate analyses. We found high genetic and phenotypic correlations among all the BW traits and the $As_{wt}$; the correlations generally increased as the age of the birds increased. Genetic and phenotypic correlations between the BW120 and $As_{wt}$ were both found close to 1, reflecting the similarity of the upper asymptote and BW at the later ages. The parameters $t_{mid}$ and scale showed a strong positive genetic and phenotypic correlation while both have negative genetic and positive phenotypic correlations with $As_{wt}$. The negative genetic correlation between $As_{wt}$ and $t_{mid}$ is considered favorable since individuals with high $As_{wt}$ will take less time to reach $t_{mid}$ making that individuals with high asymptotic weight can be identified earlier. Similarly, positive genetic correlation between $t_{mid}$ and scale is also considered favorable and logical because for birds that take less time to reach 50% of the asymptotic weight we will automatically see shrinkage in the scale. A smaller value for scale also means asymptotic weight will be approached earlier. In other studies a negative genetic correlation was also observed between $As_{wt}$ and exponential rate of decay of the specific growth rate (k) by Mignon-Grasteau et al. [17] and between $As_{wt}$ and scaling parameter by Narinic

46

et al. [43] who applied the Gompertz model in their work on chickens and quails respectively.

In our study, pHu showed highly negative genetic correlations with PDL, a* and b* whereas correlation with L* was moderately negative. These negative genetic correlations of pHu were in agreement with the previous work of Le Bihan Duval et al. on turkey and chicken [2, 5, 37]. The increase in positive genetic correlation of PDL with BW traits at later ages could be due to the increase in glycogen contents of breast muscles with age, which also had a strong negative genetic correlation with pHu [2]. The negative genetic correlation of pHu with L* and b* would explain off color meat (PSE) with low pHu and high drip loss and vice versa which was in agreement with the results from previous studies [6, 7].

In our study, both the PDL and PBM were recorded in percentages, and the genetic and phenotypic correlations between these traits were close to zero. We found that PBM had positive genetic and phenotypic correlations with BrL and BrW. The high genetic and phenotypic correlation between BrL and BW traits was also observed by Adeyinka et al. [28] in chickens. The positive genetic and phenotypic correlation of PBM with BrL and BrW will be useful in selection for increased PBM which is an important trait but can only be recorded after the animal is killed.

## 2.5 Conclusion

The results of this analysis, in particular the correlations between weights as well as the growth curve traits (Additional file 2.1), suggest that the turkey birds could be selected for breeding at earlier time points, between 60 and 80 days of age, in order to improve overall production and the yield of desirable cuts of meat at slaughter age. The selection of birds within this age range for high BW would also increase growth rates. Attention would need to be given to meat quality traits, drip loss and pHu which had low heritabilities but quality of meat would still be expected to deteriorate from selection on early body weight.

## Authors' contributions

MG, RPMAC and AV designed the study, MLA, JB and BD analyzed the data. MLA and JB wrote the manuscript and all other authors gave suggestions and comments for the improvement. All authors read and approved the final manuscript.

## Acknowledgements

## Additional files

### *Additional file 2.1*

**Link:** http://www.biomedcentral.com/1471-2156/12/14/additional

**Description**

**Title: Estimated genetic parameters (heritabilities and correlations with standard errors) for different traits in turkey birds.** Heritabilities (diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) are presented with standard errors (in parenthesis) for the different traits. BW01, BW17, BW40, BW60, BW80, and BW120 are the BW at days 1,17, 40, 60, 80, and 120 of age; PBM = percentage breast meat at 20 wk of age; BrL = breast length at 20 wk of age; BrW = breast width at 20 wk of age; PDL = percent drip loss at 20 wk of age; pHu = ultimate pH at 20 wk of age; L* = lightness at 20 wk of age; a* = redness at 20 wk of age; b* = yellowness at 20 wk of age; $As_{wt}$ = upper asymptote (estimated growth curve parameter); $t_{mid}$ = inflection point at 50% asymptote (estimated growth curve parameter); scale = constant that is proportional to the overall growth rate (estimated growth curve parameter).

# References

1.  Windhorst HW: **Changing regional patterns of turkey production and turkey meat trade**. *Worlds Poult Sci J* 2006, **62**(01):97-114.
2.  Le Bihan-Duval E, Debut M, Berri CM, Sellier N, Santé-Lhoutellier V, Jégo Y, Beaumont C: **Chicken meat quality : genetic variability and relationship with growth and muscle characteristics**. *BMC Genetics* 2008, **9**:53.
3.  Berri C: **Effect of selection for improved body composition on muscle and meat characteristics of broilers from experimental and commercial lines**. *Poult Sci* 2001, **80**:833-838.
4.  Li H, Deeb N, Zhou H, Mitchell AD, Ashwell CM, Lamont SJ: **Chicken quantitative trait loci for growth and body composition associated with transforming growth factor-β genes1**. *Poult Sci* 2003, **82**:347-356.
5.  Le Bihan-Duval E, Berri C, Baeza E, Millet N, Beaumont C: **Estimation of the genetic parameters of meat characteristics and of their genetic correlations with growth and body composition in an experimental broiler line**. *Poult Sci* 2001, **80**:839-843.
6.  Barbut S: **Problem of pale soft exudative meat in broiler chickens.** *Br Poult Sci* 1997, **38**:355-358.
7.  Kijowski J, Niewiarowicz A: **Emulsifying properties of proteins and meat from broiler breast muscles as affected by their initial pH values**. *J Food Technol* 1978, **13**:451-459.
8.  Allen CD, Russel SM, Fletcher DL: **The relationship of broiler breast meat color and pH to shelf-life and odor development**. *Poult Sci* 1997, **76**:1042-1046.
9.  Liu G, Dunnington EA, Siegel PB: **Maternal effects and heterosis for growth in reciprocal cross population of chickens**. *J Anim Breed Genet* 1993, **110**:423-428.
10. Nestor KE, McCartney MG, Harvey WR: **Genetics of growth and reproduction in the turkey. 1. Genetic and non-genetic variation in body weight and body measurements**. *Poult Sci* 1967, **46**:1374-1384.
11. Arthur JA, Abplanalp H: **Linear estimates of heritability and genetic correlation for egg production, body weight, conformation and egg weight of turkeys**. *Poult Sci* 1975, **54**:11-23.
12. Kranis A, Hocking PM, Hill WG, Woolliams JA: **Genetic parameters for a heavy female turkey line: impact of simultaneous selection for body weight and total egg number**. *Br Poult Sci* 2006, **47**(6):685 - 693.

13. Nestor KE, Anderson JW, Patterson RA: **Genetics of growth and reproduction in the turkey. 14. Changes in genetic parameters over thirty generations of selection for increased body weight**. *Poult Sci* 2000, **79**(4):445-452.

14. Toelle VD, Havenstein GB, Nestor KE, Bacon WL: **Estimates of genetic parameters in turkeys. 3. Sexual dimorphism and its implication in selection procedures**. *Poult Sci* 1990, **69**:1634–1643.

15. Fitzhugh HA: **Analysis of growth curves and strategies for altering their shape**. *J Anim Sci* 1976, **42**(4):1036 1051.

16. Sengul T, Kuraz S: **Non-Linear Models for Growth Curves in Large White Turkeys**. *Turk J Vet Anim Sci* 2005, **29**:331-337.

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

18. Koivula M, Sevon-Aimonen M-L, Stranden I, Matilainen K, Serenius T, Stalder KJ, Mantysaari EA: **Genetic (co)variances and breeding value estimation of Gompertz growth curve parameters in Finnish Yorkshire boars, gilts and barrows**. *J Anim Breed Genet* 2008, **125**:168-175.

19. Coster A: **Pedigree: Pedigree 2009.** [R package version 1.2].

20. Honikel KO: **Reference methods for the assessment of physical characteristics of meat**. *Meat Sci* 1998, **49**:447-457.

21. Ihaka R, R. Gentleman: **"R: A language for data analysis and graphics**. *J Comput Graph Stat* 1996, **5**(5):299-314.

22. Ricklefs RE: **Pattern of growth in birds**. *Ibis* 1968a, **110**:419-451.

23. Zach R, Mayoh KR: **Weight and feather growth of nestling tree swallows**. *Can J Zool* 1982, **60**:1080-1090.

24. SAS Institute: **SAS User's Guide: Statistics**. In.: SAS Institute Inc., Cary, NC; 2000.

25. Gilmour AR, Gogel BJ, Cullis BR, Thompson R: **ASReml User Guide Release 2.0** In.: VSN International Ltd Hemel Hempstead HP1 1ES, UK; 2006.

26. Chapuis H, Tixer-Boichard M, Delabrosse Y, Ducroc V: **Multivariate restricted maximum likelihood estimation of genetic parameters for production traits in three selected turkey strains**. *Genet Sel Evol* 1996, **28**:299–317.

27. Nestor KE, Anderson JW, Patterson RA, Velleman SG: **Genetics of growth and reproduction in the turkey. 17. changes in genetic parameters over**

**forty generations of selection for increased sixteen-week body weight**. *Poult Sci* 2008, **87**:1971-1979.

28. Adeyinka IA, Oni OO, Nwagu BI, Adeyinka FD: **Genetic Parameter Estimates of Body Weights of Naked Neck Broiler Chickens**. *Int J Poult Sci* 2006, **5**(6):589-592.

29. Zerehdaran S, Vereijken ALJ, Arendonk JAMv, Waaij EHvd: **Estimation of Genetic Parameters for Fat Deposition and Carcass Traits in Broilers**. *Poult Sci* 2004, **83**:521–525.

30. Miraei-Ashtiani SR, Seyedalian SAR, Shahrbabak MM: **Variance components and heritabilities for body weight traits in Sangsari sheep, using univariate and multivariate animal models**. *Small Rumin Res* 2007, **73**:109-114.

31. Buss EG: **Genetics of growth and meat production in turkeys, in: CRAWFORD, R.D. (Ed.)** *Poultry Breeding and Genetics* 1990: 645—675(Amsterdam, Elsevier).

32. Chambers JR: **Genetics of growth and meat production in chicken, in: CRAWFORD, R.D. (Ed.)** *Poultry Breeding and Genetics* 1990:559 (Amsterdam, Elsevier).

33. Le Bihan-Duval E, Mignon-Grasteau S, Millet N, Beaumont C: **Genetic analysis of a selection experiment on increased body weight and breast muscle weight as well as on limited abdominal fat weight**. *Br Poult Sci* 1998, **39**:346–353.

34. Koerhuis ANM: **Non-normality of egg production distributions in poultry and the effects of outlier elimination and transformation on size and curvilinearity of heritability**. *Livest Prod Sci* 1996, **45**:69-85.

35. Tullett SG, Burton GF: **Factors affecting the weight and water status of the chick at hatch**. *Br Poult Sci* 1982, **23**:361-369.

36. North M: **Manual de Production Avicola**, vol. 3, Edition editorial El Manual Moderno edn. Mexico; 1986.

37. Le Bihan-Duval E, Berri C, Baéza É, Santé VA, T. , Rémignon H, Le Pottier G, Bentley J, Beaumont C, Fernandez X: **Genetic parameters of meat technological quality traits in a grand-parental commercial line of turkey**. *Genet Sel Evol* 2003, **35** 623-635.

38. Richards FJ: **A flexible growth function for empirical use**. *J Exp Botany* 1959, **10**:290-300.

39. Aggrey SE: **Comparison of three nonlinear and Spline regression models for describing chicken growth curves**. *Poult Sci* 2002 **81**:1782-1788.

51

40.     Grossman M, Bohren BB: **Logistic growth curve of chickens: heritability of parameters**. *J Hered* 1985, **76**:459-462.

41.     Le Rouzic A, Alvarez-Castro JM, Carlborg O: **Dissection of the genetic architecture of body weight in chicken reveals the impact of epistasis on domestication traits**. *Genetics* 2008, **179**(3):1591-1599.

42.     Barbato GF, Vasilatos-Younken R: **Sex-linked and maternal effects on growth in chickens**. *Poult Sci* 1991, **70**:709-718.

43.     Narinc D, Aksoy T, Karaman E: **Genetic parameters of growth curve parameters and weekly body weights of Japanese Quails (Coturnix coturnix japonica)**. *J Ani Vet Adv* 2010, **9**(3):501-507.

# 3

# A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the Turkey and Chicken genomes

Muhammad L Aslam[1], John WM Bastiaansen[1], Richard PMA Crooijmans[1], Addie Vereijken[2], Hendrik-Jan Megens[1], Martien AM Groenen[1]

[1] Animal Breeding and Genomics Centre, Wageningen University,6709PG, Wageningen, The Netherlands
[2] Hendrix Genetics, Research & Technology Centre, 5830 AC, Boxmeer, The Netherlands

## Abstract

### Background

The turkey (*Meleagris gallopavo*) is an important agricultural species that is the second largest contributor to the world's poultry meat production. The genomic resources of turkey provide turkey breeders with tools needed for the genetic improvement of commercial breeds of turkey for economically important traits. A linkage map of turkey is essential not only for the mapping of quantitative trait loci, but also as a framework to enable the assignment of sequence contigs to specific chromosomes. Comparative genomics with chicken provides insight into mechanisms of genome evolution and helps in identifying rare genomic events such as genomic rearrangements and duplications/deletions.

### Results

Eighteen full sib families, comprising 1008 (35 F1 and 973 F2) birds, were genotyped for 775 single nucleotide polymorphisms (SNPs). Of the 775 SNPs, 570 were informative and used to construct a linkage map in turkey. The final map contains 531 markers in 28 linkage groups. The total genetic distance covered by these linkage groups is 2,324 centimorgans (cM) with the largest linkage group (81 loci) measuring 326 cM. Average marker interval for all markers across the 28 linkage groups is 4.6 cM. Comparative mapping of turkey and chicken revealed two inter-, and 57 intrachromosomal rearrangements between these two species.

### Conclusion

Our turkey genetic map of 531 markers reveals a genome length of 2,324 cM. Our linkage map provides an improvement of previously published maps because of the more even distribution of the markers and because the map is completely based on SNP markers enabling easier and faster genotyping assays than the microsatellite markers used in previous linkage maps. Turkey and chicken are shown to have a highly conserved genomic structure with a relatively low number of inter-, and intrachromosomal rearrangements.

## 3.1 Background

The turkey (*Meleagris gallopavo*, MGA) is an important agricultural species that is largely used as a meat type bird. In 2008, turkey represented 6.65% of the world poultry meat production [1]. The world-wide turkey population has rapidly grown due to increased commercial farming. Global turkey stocks nearly tripled from 178 million in 1970 to over 482 million in 2008. Over the same time period, the production volume increased more than fivefold from 1.2 to 6.1 million tons [1].

The turkey genome consists of 39 pairs of autosomes and 1 pair of sex chromosomes [2]. The predicted size of the turkey genome is 1.1 billion bases on the turkey genome build UMD 2.01 which is based on sequences from a combination of two next generation sequencing platforms, Roche 454 and Illumina GAII with 5X and 25X coverage respectively. Limited information is available on the turkey linkage and physical map although a small number of low resolution linkage maps using microsatellite markers [2, 3] have been published. Linkage maps of chicken on the other hand are more abundant and have generally used larger numbers of markers [4-10]. Comparative cytogenetic and linkage maps between turkey and chicken showed conserved synteny and close ancestral relation among these species [2, 3] and support the hypothetical ancestral Galliform karyotype [11]. Chromosome banding and zoo-FISH with chromosome paints for the turkey and chicken chromosomes have suggested that chicken and turkey karyotypes are distinguished by at least two interchromosomal rearrangements [2, 12, 13]. Chicken chromosome 2 and 4 are represented by turkey chromosomes 3 and 6 and by turkey chromosomes 4 and 9 respectively [2, 3, 13].

Chromosome studies have revealed that the karyotype is more conserved among avian lineages than it is among other groups, such as mammals, with most avian species showing a diploid chromosome number between 76 and 80 (http://www.genomesize.com). This suggests that chromosomal evolution or large-scale rearrangements affecting chromosome number occur at a low rate in birds and as a result many chromosomes have remained more or less intact during avian evolution [14]. Chicken chromosome specific probes have been used for in situ hybridization onto metaphase spreads of other birds and revealed an overall picture of a high degree of chromosomal homology between chicken and representatives from many avian orders [15]. Hybridization results also indicated that interchromosomal rearrangements have been rare during avian evolution [16, 17].

A linkage map is essential for the mapping of quantitative trait loci (QTL) and very useful for the assembly of genome sequence and subsequently mapping of genes along the chromosomes. A high-resolution linkage map facilitates fine mapping of

quantitative trait loci (QTLs) and can be produced because of the abundance of SNPs within the genome [18]. SNP based genotyping is preferred because it is highly accurate, quick and automated, using limited human intervention. Increasing the marker density of the linkage map further enables the analyses of genomic sequences associated with high recombination rates [9].

The present study was designed to develop a SNP based linkage map in turkey and to detect genomic rearrangements between turkey and chicken.

## 3.2 Methods

### Experimental population

Parents were randomly selected from two different lines to produce F1 offspring. Ten parent males were randomly selected from a line that was selected for high growth and ten parent females were randomly selected from a line that was selected for high reproduction. Average body weight of males in the high growth line from which ten parent males were randomly selected was 20.6 Kg at 20 weeks of age and the average egg production of females in high reproduction line from which ten parent females were randomly selected was 115.5 hatching eggs/24 weeks. An F2 generation of 18 full sib families was produced by crossing 17 randomly selected F1 males and 18 randomly selected F1 females. One male was mated with two females, other F1 parents were mated only once. In total, 973 F2 offspring were produced with an average full sib family size of 54.1 with a range from 31-90 individuals. All families were used for the SNPs genotyping to construct linkage maps of different chromosomes.

### DNA isolation

Genomic DNA was isolated from blood samples collected in 10% EDTA using either the automated nucleic acid extraction CAS-1820 X-tractor Gene (Corbett Life Science), or the manual nucleic acid extraction using Gentra Puregene Blood Kit (Qiagen) following manufacturer's protocol with minor modifications.

DNA concentrations were measured using ND-1000 Spectrophotometer (NanoDrop) and diluted to the required concentration of 50 ηg $\mu L^{-1}$.

### SNP selection

Previously, we identified 11,287 SNPs in turkey by sequencing reduced representation libraries on an Illumina GA sequencer [19]. To achieve an even spacing of SNPs across the 40 turkey chromosomes while a turkey genome sequence was not available, SNPs in turkey were selected based on their orthologous position on the chicken genome sequence (WASHUC2 build, May 2006). Currently the chicken genome [20] covers 30 of the 39 chromosomes in chicken which comprises approximately 95 % of chicken genome. By this approach,

we did not select SNPs in parts of the turkey genome that are syntenic to genomic regions in chicken that are currently not represented in the chicken genome assembly. Assembled turkey short read contigs from Kerstens et al. [19] that contained SNPs were mapped on the chicken genome. Short read contigs in the size range of 50-100 bp were mapped using Megablast [21] and short read contigs of 100 bp and longer were mapped using BlastZ [22] with contig alignment criteria of at least 80% alignment and at least 60% sequence identity. In total 6,537 SNPs could be assigned a syntenic location on the chicken genome. In addition to chicken genome location, the final selection criteria for SNPs also included the Illumina design score and the estimated minor allele frequency based on the Illumina sequences from Kerstens et al [19]. The distance (in bps) between the selected SNPs was varied based on the size of the chromosome, because of the higher recombination frequency on the microchromosomes of birds. Chicken chromosomes were divided into three groups; 1-10 + Z, 11-19 and 20-28 + LGE22 and the average SNP spacing chosen for these three groups was 1.4-1.9 SNP per Mb, 0.7-1.0 SNP per Mb and 0.4-0.6 SNP per Mb respectively.

In addition, seven SNPs derived from 5 different turkey genes i.e. *Pit1*, *AFABP*, *PRKAG3*, *IGF2* and *GDF8* were also used.

### *Genotyping*

Two 384-plex GoldenGate oligo pool assay (OPA) sets were designed for genotyping using VeraCode technology on an Illumina BeadXpress Reader. The GoldenGate assay was performed according to manufacturer's protocol and as described in Fan et al. [23] and Hyten et al. [24]. Automated genotype clustering and calling was performed with GenomeStudio™ data analysis software (Illumina). All genotype calling results were manually checked and any obvious errors in calling the homozygous or heterozygous clusters were corrected.

SNPs selected from the 5 turkey genes (*Pit1*, *AFABP*, *PRKAG3*, *IGF2* and *GDF8*) were genotyped with an ABI SNaPshot assay and analyzed on an ABI 3730 DNA Analyzer (Additional file 3.1).

### *Genetic Linkage analysis*

Genotyping data was filtered by removing uninformative markers, markers giving Mendelian errors in more than one families and markers with low call rate as described by Groenen et al. [9]. The modified CRI-MAP software version 2.4 [25] by Xuelu Liu (Monsanto), which can handle much larger numbers of markers segregating in complex pedigrees was utilized for the linkage analysis.

Map building was performed step by step using AUTOGROUP, BUILD, CHROMPIC, FLIPSN, and FIXED options of CRI-MAP according to the procedures used by Stapley et al. [26] and Elferink et al. [10]. Using AUTOGROUP, parameter layers utilized for

57

getting linkage groups were as follows: layer 1 (20, 0, 2, 0.3); layer 2 (20, 0, 20, 0.3); layer 3 (10, 0, 20, 0.3) and layer 4 (5, 0, 20, 0.3). Layer 4 had minimum stringency with likelihood ratio (LOD score) >5, 0 times the average number of meiosis, shared linkages with not more than 20 groups and with 0.3 of minimum linkage ratio [25]. Linkage groups were assigned to specific turkey chromosomes using the already known physical positions of turkey SNPs in the chicken genome and comparative information from the cytogenetic study of Griffin et al. [2] on turkey and chicken. Turkey chromosome names were assigned using the nomenclature used by Griffin et al. [2].

Maps are reported as sex averaged maps unless otherwise indicated and map figures were drawn with the MapChart software version 2.2 [27].

### *Comparative genetic analysis*

The order of SNPs on our linkage map was compared to the expected order based on the turkey and chicken genome assemblies UMD 2.01 and WASHUC2, respectively. Positions on the chicken genome were obtained earlier in the SNP selection step. Positions on the turkey genome were obtained by aligning SNP flanking sequences ($<1.0 \times E^{-4}$) using BLAST with megablast option [28] against the turkey reference genome sequence (UMD 2.01).

The turkey physical map order of SNPs was used to validate the linkage map order with CRI-MAP using the BUILD option. The order of SNPs in linkage maps was modified if the physical map order had a higher likelihood and total chromosome map length was smaller than the linkage map order. The genetic distance between the terminal markers of every chromosome from the turkey linkage map was compared to the genetic distance between the corresponding positions of the chicken genome. First, the sequence positions (bp) of these terminal turkey markers were found on the chicken physical map. Second, chicken markers were taken from the study of Elferink et al. [10] at the closest position (bp) to these sequence positions (bp). Finally the genetic distance between these chicken markers was calculated and compared to the turkey map length.

### *Analysis of recombination rate and sequence motif densities*

The physical distance (Mb) on turkey chromosomes was calculated between the first and the last SNP of the linkage map using the blastall option in BLAST [28]. Number of Mb covered by the linkage map (cM) was used to calculate recombination rate (cM/Mb) for every turkey chromosome which was compared to the physical size (Mb) of the chromosomes [9, 26]. The recombination rates (cM/Mb) were also compared to those for the chicken chromosomes described by Elferink et al. [10].

58

Densities of sequence motifs/elements CCCCCCC, CCTCCCT, CTCTCCC, CpG and CTCF consensus sequence CCNCCNGGNGG were found to vary with chromosome in chicken [9], therefore we also calculated these densities for each turkey chromosome from the turkey genome sequence (UMD 2.01). Only the part of the chromosome sequence covered by the linkage map was used to calculate these densities. Number of elements per Mb was calculated and compared against chromosome length (cM) except for CpG that was compared against cM/Mb [9].

### *Ethical approval for the use of animals in this study*

Although animals were used in this experimental work, no direct experiments were performed on them. Blood sample collection was carried out by licensed and authorized personnel under approval of Hendrix Genetics. No approval from the ethics committee was necessary.


## 3.3 Results

### *Genotyping results*

Genotyping call rates with an average of 0.80 were obtained. In total, 775 SNPs (2x 384-plex GoldenGate + 7 additional SNPs) were selected for genotyping and out of these, 98 SNP assays failed (missing genotypes in the whole population), 80 SNPs appeared to be monomorphic (AA, or BB genotype) or positive for parologous sequences (all genotypes AB), 13 SNPs showed non-Mendelian inheritance in more than one family and 14 SNPs had zero informative meiosis. In total 205 SNPs were removed from the dataset.

### *Linkage maps*

After filtering of genotyping data, 570 SNP markers were left for the linkage analysis. Of the total 570 markers that met all quality criteria, 531 markers were found significantly linked which were subsequently inserted at their most likely position (BUILD option, LOD > 3) on one of 28 linkage groups that subsequently were assigned to 27 autosomes and the Z chromosome (Table 3.1). The number of informative meiosis for a marker varied from 7 to 666 with an average of 255. The largest chromosome, MGA1, had a map with 81 SNPs and a map size of 325.8 cM, followed by MGA2 with 55 SNPs and a map size of 229 cM. The chromosomes MGA25 and MGA30 had the lowest number of SNPs (4 each) as well as the smallest map sizes with map lengths of 23.5 and 6.3 cM respectively (Table 3.1). The total length of the sex average map (excluding the Z chromosome) was 2,165 cM and the average marker spacing was 4.4 cM. Sex specific analysis showed a difference in the male and the female maps. For 70 % of chromosomes, male maps were longer than female maps, except for chromosomes MGA10, 11, 15, 16, 23, 25, 26 and

MGA28 where the female maps were longer (Table 3.1). In general, a difference in length of 9% was observed between sex specific maps.

**Table 3.1.** Comparison of maps of turkey and chicken chromosomes based on genetic and physical sizes.

| | Turkey | | | | | Chicken | | |
| Chr | Number of SNPs | Female (cM) | Male (cM) | Average (cM) | Length (Mb) | Chr | Genetic length (cM) | Syntenic region (Mb) |
|---|---|---|---|---|---|---|---|---|
| MGA1 | 81 | 318.9 | 344.7 | 325.8 | 200.7 | GGA1 | 353.4 | 193.3 |
| MGA2 | 55 | 216.4 | 249.9 | 229 | 115.1 | GGA3 | 233.5 | 111.7 |
| MGA3 | 40 | 140.6 | 149.5 | 140.9 | 89.7 | GGA2q | 149.5 | 134.7 |
| MGA4 | 27 | 94 | 146.7 | 120.9 | 67.4 | GGA4q | 132.1 | 70.0 |
| MGA5 | 33 | 108.1 | 131.3 | 118.3 | 59.8 | GGA5 | 116.2 | 61.2 |
| MGA6 | 23 | 98.9 | 111.9 | 104.4 | 48.2 | GGA2p | 111.9 | 131.7 |
| MGA7 | 22 | 65.3 | 77.3 | 71.4 | 29.6 | GGA7 | 102.5 | 37.5 |
| MGA8 | 19 | 64.1 | 67.4 | 67.3 | 32.2 | GGA6 | 82.5 | 33.0 |
| MGA9 | 10 | 52.9 | 59.7 | 55.2 | 16.8 | GGA4p | 60.5 | 91.3 |
| MGA10 | 21 | 82.1 | 65.7 | 76.7 | 30.0 | GGA8 | 56.0 | 29.7 |
| MGA11 | 19 | 64.1 | 54.6 | 59.8 | 22.8 | GGA9 | 78.0 | 22.3 |
| MGA12 | 14 | 56.6 | 63.1 | 58.4 | 14.1 | GGA10 | 45.6 | 18.8 |
| MGA13 | 17 | 51.2 | 59.8 | 54.4 | 18.0 | GGA11 | 62.9 | 20.9 |
| MGA14 | 13 | 50.7 | 59.9 | 55.1 | 14.0 | GGA12 | 44.3 | 13.7 |
| MGA15 | 21 | 59 | 56.9 | 59 | 16.1 | GGA13 | 56.2 | 17.4 |
| MGA16 | 13 | 46.7 | 41.1 | 40.7 | 12.0 | GGA14 | 47.5 | 12.1 |
| MGA17 | 13 | 57.5 | 59.5 | 57.5 | 12.5 | GGA15 | 52.7 | 12.3 |
| MGA19 | 10 | 49.9 | 57.6 | 51.2 | 8.9 | GGA17 | 48.7 | 10.0 |
| MGA20 | 12 | 54.6 | 67.3 | 60.6 | 9.3 | GGA18 | 48.7 | 9.3 |
| MGA21 | 12 | 54.5 | 76 | 60.8 | 8.9 | GGA19 | 41.9 | 8.3 |
| MGA22 | 10 | 53.8 | 60.3 | 56.2 | 11.3 | GGA20 | 42.0 | 10.6 |
| MGA23 | 9 | 53.2 | 50.6 | 61.4 | 4.5 | GGA21 | 41.1 | 4.9 |
| MGA24 | 4 | 26 | 38.6 | 33.1 | 1.9 | GGA22 | 21.6 | 1.8 |
| MGA25 | 4 | 25.2 | 22.2 | 23.5 | 4.3 | GGA23 | 30.8 | 4.8 |
| MGA26 | 8 | 71.3 | 50.7 | 57.3 | 6.0 | GGA24 | 51.8 | 5.7 |
| MGA28 | 8 | 60.2 | 45 | 52.6 | 4.2 | GGA26 | 45.5 | 4.3 |
| MGA30 | 4 | 1.4 | 4.9 | 6.3 | 1.1 | GGA28 | 17.1 | 0.9 |
| **Total autosomal** | 522 | 2077.2 | 2272.2 | 2164.8 | 859.4 | **Total** | 2174.5 | 1072.2 |
| MGAZ | 9 | --- | 159.1 | 159.1 | 80.02 | GGAZ | 221.9 | 74.3 |
| **Total** | **531** | **2077.2** | **2431.3** | **2323.9** | **939.4** | **Total** | **2396.4** | **1146.5** |

### *Comparative genetic results*

*Marker order:* For all except three of the turkey chromosomes the comparison of the linkage and the physical maps did not reveal any differences. For the three chromosomes, MGA2, 11 and 17, the marker order from the physical maps, showed a higher likelihood and a smaller map distance than the marker order obtained from our linkage analyses. Log likelihood values for MGA2, 11 and 17 were increased by 20.6, 98.6 and 0.7 and map distance reduced by 4.0, 17.0 and 1.3 cM respectively. For these three chromosomes the marker order based on the physical map was used in further analyses.

Marker orders were found to be highly conserved between the turkey linkage and the chicken physical maps although 57 rearrangements were still detected between these species. The order of the SNP markers on chromosomes MGA14, 21, 25, 26 and MGAZ even showed 100 % accordance with the order in the syntenic chicken chromosomes (Additional file 3.2).

The linkage maps for the turkey and the chicken chromosomes generally showed small differences in their lengths. Three exceptions are turkey chromosomes MGA1, MGA7 and MGAZ that showed a difference of more than 25 cM with their syntenic chicken chromosomes GGA1 GGA7 and GGAZ. Whole genome genetic map size of chicken was 72.5 cM larger than the whole genome genetic map size of turkey. In the comparisons of genetic lengths of turkey and chicken chromosomes, the difference in the reference genome positions (bp) of turkey SNPs genotyped in the present study and the genome positions (bp) of chicken SNPs used in the study by Elferink et al. [10] were small. On average the distance between the reference positions was 58,614 bp which will have caused an average difference of 0.28 cM/chromosome based on the average figure of 4.8 cM per Mb in Turkey. The total physical map size of turkey covered by markers genotyped in this study was 939.4 Mb. This is smaller than the region of the chicken physical map covered by the turkey genetic map which is 1146.5 Mb (Table 3.1).

*Rearrangements:* Two interchromosomal and 57 intrachromosomal rearrangements were observed between turkey and chicken (Figure 3.1). Two linkage groups, MGA3 and MGA6 were obtained from the SNPs selected with syntenic positions on chicken chromosome 2 and similarly two linkage groups, MGA4 and 9 were obtained from the SNPs selected from chicken chromosome 4 (Figure 3.1). These chromosomes (MGA3, 6 and MGA4, 9 Vs GGA2 and GGA3 respectively) did not only show interchromosomal rearrangements, but also showed multiple intrachromosomal rearrangements between turkey and chicken (Figure 3.1).
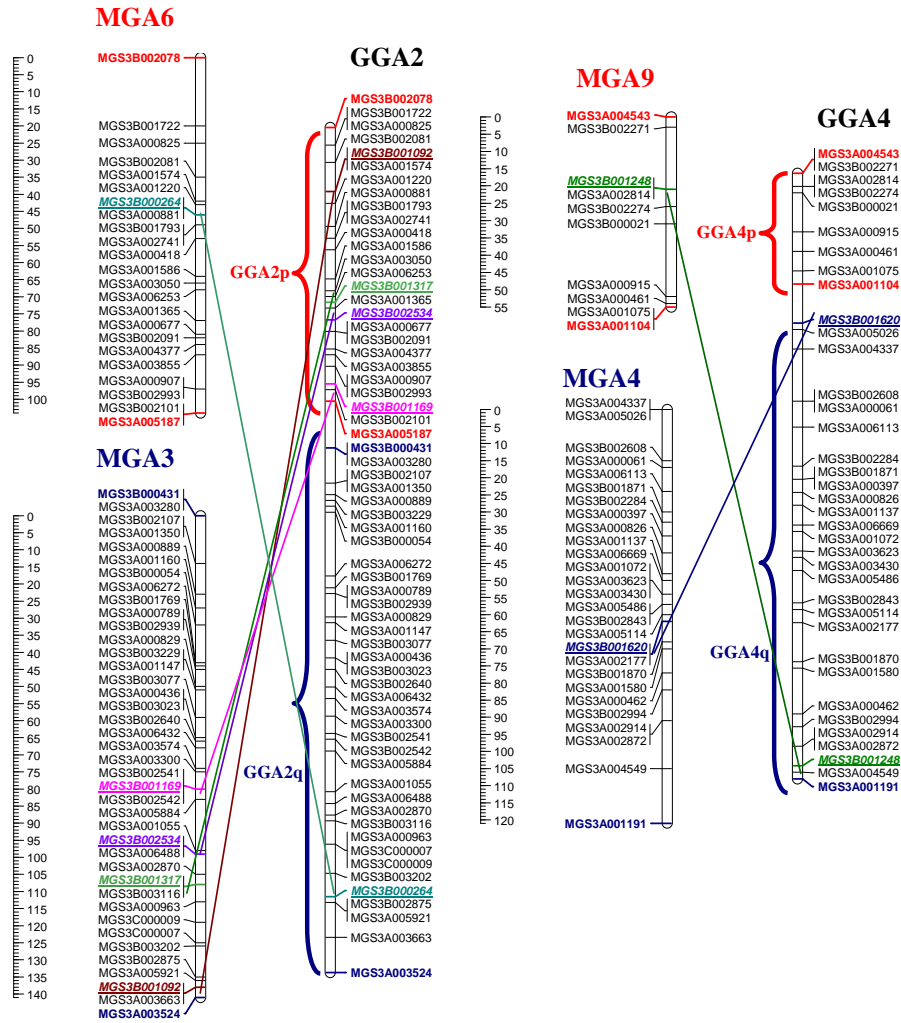
**Figure 3.1.** Inter and intra chromosomal rearrangements between turkey and chicken. Linkage maps of turkey chromosomes MGA3, MGA6 and MGA4, MGA9 showing inter and intra chromosomal rearrangements (fission, fusion and inversions,) with their syntenic chicken chromosomes GGA2 and GGA4 (maps based on physical position of SNPs in chicken genome).

Regions with inverted marker order were observed on turkey chromosomes 10 and 20 when compared to their syntenic chicken chromosomes GGA8 and GGA18 (Figure 3.2). Other complex intrachromosomal rearrangements were also observed

on turkey chromosome 1, 2, 5, 8, 11, 12, 13, 15, 16, 17, 19, 22, and 28 when compared to their syntenic chicken chromosomes (Additional Files 2 & 3).

The number of rearrangements per Mb varied considerably for different chromosomes. The average number of rearrangements per Mb for larger chromosomes (MGA1-MGA10) was 0.06, ranging from 0.01-0.13 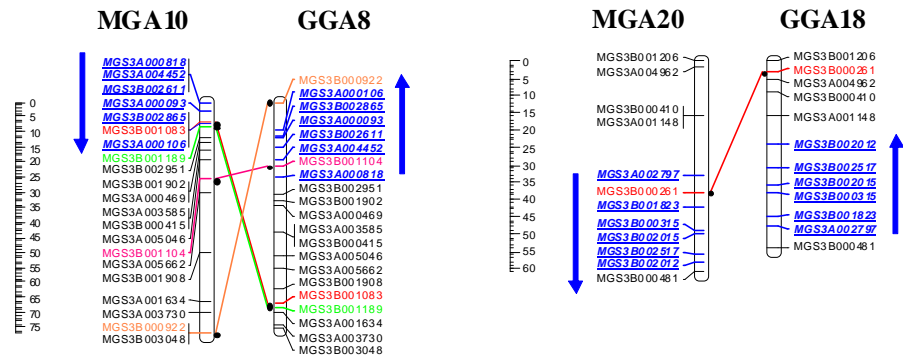with highest rate of rearrangements of 0.13 per Mb on MGA10. The average number of rearrangements per Mb for the smaller chromosomes (MGA11-MGA30) was 0.11, ranging from 0.08 - 0.42 with highest rate of rearrangements of 0.42 per Mb on MGA12.



**Figure 3.2.** Intrachromosomal rearrangements between turkey and chicken involving nearly a complete chromosome arm. Turkey chromosomes MGA10 and MGA20 (genetic linkage maps) showing intrachromosomal rearrangements (Inversions) compared to the syntenic chicken chromosomes GGA8 and GGA18 (maps based on physical position of SNPs in chicken genome).

Comparative analysis of the turkey linkage, the turkey physical and the chicken physical maps showed discordance in the chromosomal allocation of 6 SNPs to these maps (Table 3.2). The turkey linkage and the chicken physical maps agreed with each other in the chromosomal allocation of these 6 SNPs while the turkey physical map disagreed. For example, according to the turkey linkage and chicken physical maps the SNP MGS3A000968 was assigned to MGA1 and GGA1 while this SNP was positioned on MGA8 in the turkey physical map (Table 3.2). Fourteen SNPs could not be assigned to any position on the turkey physical map while the allocation of these 14 SNPs to the turkey linkage map and the chicken physical map also agreed with each other (Table 3.2).

63

**Table 3.2.** SNPs with discordance in allocation on turkey genome with turkey genetic and chicken physical map.

| SNP_ID | Turkey Linkage map (MGA) | Turkey physical map (MGA) | Chicken physical map (GGA) |
|---|---|---|---|
| MGS3A000968 | 1 | 8 | 1 |
| MGS3A003050 | 6 | 13 | 2 |
| MGS3A000053 | 8 | 19 | 6 |
| MGS3A004543 | 9 | 10 | 4 |
| MGS3A000578 | 15 | 14 | 13 |
| MGS3B002546 | 19 | 8 | 17 |
| MGS3A005799 | 1 | NA | 1 |
| MGS3B003240 | 2 | NA | 3 |
| MGS3A007335 | 2 | NA | 3 |
| MGS3A005026 | 4 | NA | 4 |
| MGS3B000939 | 5 | NA | 5 |
| MGS3A007520 | 5 | NA | 5 |
| MGS3A006539 | 7 | NA | 7 |
| MGS3A007601 | 15 | NA | 13 |
| MGS3A007553 | 16 | NA | 14 |
| MGS3A002797 | 20 | NA | 18 |
| MGS3C000006 | 1 | NA | 1 |
| MGS3C000009 | 3 | NA | 2 |
| MGS3B001450 | Z | NA | Z |
| MGS3B002754 | Z | NA | Z |

SNP chromosomal assignment by turkey linkage, turkey sequence and chicken sequence maps; NA = Not Aligned

### *Recombination rate and sequence elements*

Recombination rate of turkey chromosomes varied from 1.6 to 17.2 cM/Mb. The physical length of chromosomes showed an inverse relation with recombination rate while CpG/Mb density across the chromosome showed a direct relation. Turkey and chicken chromosomes of smaller sizes showed higher recombination rates than chromosomes with larger sizes (Figure 3.3A). CpG content showed increasing values with increasing recombination rate, i.e. higher CpG content in smaller chromosomes (Figure 3.3B). The frequency of sequence elements (CTCF, CCTCCCT, CTCTCCC and CCCCCCC) per Mb was found to be negatively correlated with the genetic size (cM) of chromosomes (Figure 3.3C-F).
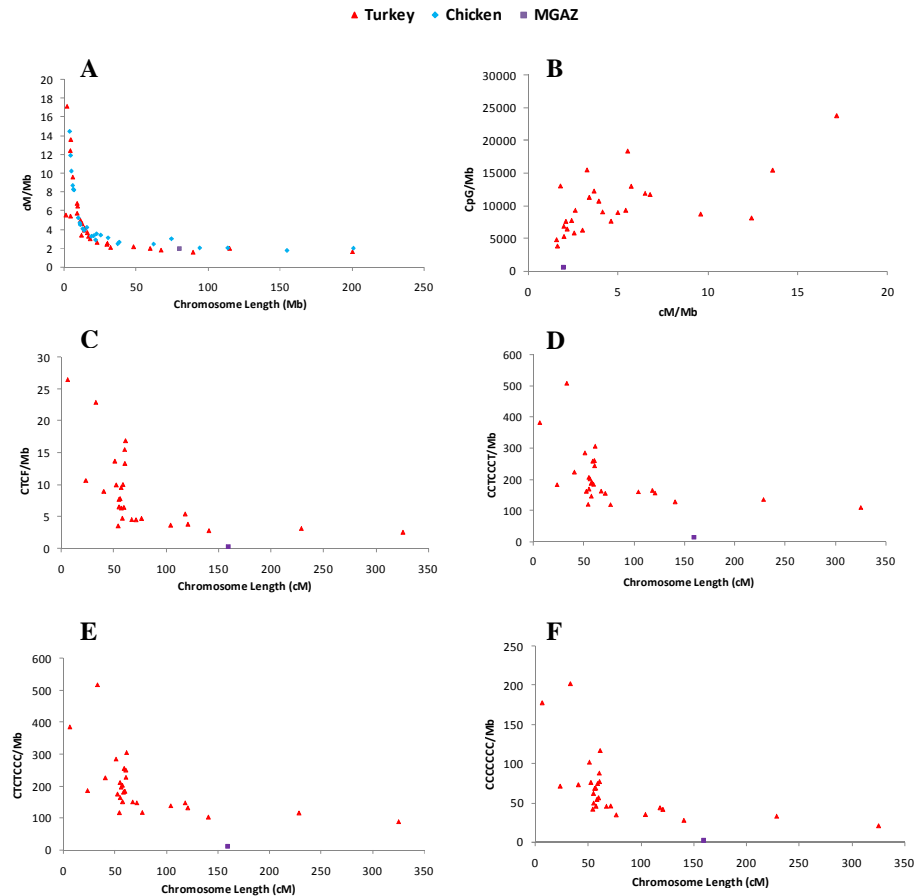
**Figure 3.3.** Distribution of recombination rates and sequence motif densities across the turkey chromosomes. Recombination rate cM/Mb was correlated with chromosome length (Mb) and CpG/Mb was correlated with recombination rate (cM/Mb). All remaining sequence motifs (CCCCCCC, CCTCCCT, CTCF and CTCTCCC) were correlated with chromosome length (cM).

## 3.4 Discussion

A whole genome SNP-based linkage map for the turkey is presented with 531 markers dispersed over 28 linkage groups and a total map length of 2324 cM. The total map length in the present study was slightly higher than that described by Reed et al. [3]. This difference in length is probably caused by coverage of three additional turkey chromosomes (MGA20, 24 and MGA25) along with the utilization of 69 % higher number of markers in the present study, likely to be covering a

larger proportion of the turkey genome. Matching the turkey linkage groups with the chicken physical map identified a map for each of the syntenic groups/chromosomes described by Griffin et al. [2]. The comparison of the syntenic chromosomes between turkey and chicken showed that the genetic lengths of turkey chromosomes were very similar to the estimated genetic length of the chicken chromosomes (Table 3.1). When comparing the turkey genetic linkage map and chicken physical map with respect to the order of markers across the chromosomes, some of the chromosomes (MGA14, 21, 25, 26 and MGAZ) showed complete conservation in the order of markers whereas others showed limited variation. The conservation in the order of markers for the chicken chromosomes GGA12, 19, 24 and GGAZ with the syntenic turkey linkage groups was also observed by Reed et al. [3]. This high rate of concordance in the order of markers between the genomes of these two avian species is indicative of a highly conserved nature of avian genomes.

Observed interchromosomal rearrangements (Figure 3.1) in the present study between turkey and chicken are in agreement with the results of Griffin et al. [2]. A number of complex intrachromosomal rearrangements (inversions) were also observed between turkey and chicken. The observed large inverted regions, of nearly a complete chromosome arm on MGA10 and MGA20 in comparison to their syntenic chicken chromosome GGA8 and GGA18 (Figure 3.2) were also observed in a sequence based comparative study by Dalloul et al. [29]. Cytogenetic studies using chromosome painting also reported an inversion on MGA10 in comparison to the syntenic chicken chromosome GGA8 [2, 13]. Our comparative linkage map of turkey and chicken does not show pericentic inversions on MGA2 and MGA3p as were reported by Griffin et al. [2] but we have observed complex rearrangements resulting in a reversed order of markers on these chromosomes (Additional file 3.3). Several other chromosomes, notably MGA1, 2, 5, 8, 11, 12, 13, 15, 16, 17, 19, 22, and MGA28 as well as the chromosomes that showed interchromosomal rearrangements (MGA3 and MGA6; MGA4 and MGA9) between turkey and chicken, also showed additional complex rearrangements probably involving multiple inversions or other complex rearrangements (Figure 3.1). A higher number of rearrangements per Mb were observed on the microchromosomes than on the macrochromosomes. The occurrence of this high number of rearrangements at the microchromosomes could be explained by the positive association of rearrangements with recombination rate [30].

Our observed low number of interchromosomal rearrangements between the chicken and turkey genomes, confirms previous results of a high degree of interchromosomal synteny in birds as seen within a number of different

66

comparative studies of chicken with quail, duck and zebra finch [30-32]. It has been suggested that the low number of interchromosomal rearrangements during avian genome evolution is a consequence of the small amount of interspersed repeats, segmental duplications, and pseudogenes in avian genomes, which provide little opportunity for non allelic homozygous recombination [33, 34]. A relatively high number of intrachromosomal rearrangements was observed in our comparative analysis of the turkey and chicken, which agrees with the findings of the sequence based comparative studies of chicken with turkey and zebra finch [26, 29, 30]. The relatively high number of intrachromosomal rearrangements clearly suggests that the organization of avian genomes is more prone to intrachromosomal rearrangements than previously appreciated based on chromosome banding and chromosome painting data [2].

The comparison of male vs. female maps showed differences in genetic lengths of maps. In turkey, the total male-specific map appeared to be 195 cM longer than the female specific map. However, female-specific maps for some chromosomes (MGA10, 11, 15, 16, 23, 25, 26 and MGA28) were also found to be longer than the male maps (Table 3.1). The longer map length in homogametic males can be explained by the Haldane-Huxley rule [35, 36], which predicts that the frequency of recombination during meiosis is lower in the heterogametic sex . The smaller map lengths in turkey for some male-specific maps were found to be an exception to the Haldane-Huxley rule. However, the longer map lengths for some chromosome maps in the heterogametic sex were also found in chicken [9].

In the present study three maps i.e. the turkey genetic linkage map, the turkey physical map and the chicken physical map were compared. The discordance of turkey physical map with the turkey genetic linkage and the chicken physical map in the allocation of marker at different chromosomes could possibly be explained by the occurrence of assembly errors in the turkey genome sequence. The turkey physical map was created completely by whole genome shotgun sequencing using Roche 454 and Illumina GA2 sequence data. Inconsistencies between the turkey linkage and chicken physical maps relative to the turkey physical map are most likely a reflection of the challenge of correctly assembling a genome based on next-gen sequencing data alone. Markers that were in agreement between turkey linkage and chicken physical maps but that could not be positioned on the turkey physical map most likely reflect an uncovered genomic regions since the turkey genome sequence is known to cover around 95% of the complete genome (Turkey genome build UMD 2.01).

In general, higher recombination rates and higher densities of GC-rich elements were found on microchromosomes compared to macrochromosomes (Figure 3.3A

67

& B). During meiosis, at least one chiasma per bivalent chromosome is required [37], but the likelihood of chiasmata forming varies along the chromosome [38]. In turkey, recombination rate and GC rich sequences (CTCF, CCTCCCT, CTCTCCC and CCCCCCC) were found to co-vary among different chromosomes. A similar trend was also seen in human, mouse and other birds like chicken and zebra finch [9, 10, 20, 26, 37, 39, 40].

In the present study recombination rates were found to be correlated with CpG/Mb. In general CpG/Mb tended to increase in areas of higher recombination i.e. microchromosomes (Figure 3.3B). This demonstrates that in the turkey microchromosomes, high recombination rate, high amount of GC-rich sequences (CTCF, CCTCCCT, CTCTCCC and CCCCCCC) and high amount of CpG contents are all correlated (Figure 3.3A-F). Other studies reported that GC-rich regions in a genome had higher gene densities [41, 42] and that microchromosomes had higher gene densities than the macrochromosomes [38]. The nature of the microchromosomes in birds, with their high recombination rates, high amount of GC-rich sequences, GC content and gene densities appears to be an extreme instance of a general trend.

The results for MGAZ in the analysis of recombination rate and sequence motif densities across the chromosomes, were unexpected and MGAZ appeared as outlier as seen in Figure 3.3A-F. This outlier spot could represent a true characteristic of MGAZ but more likely results from the low marker density on this particular chromosome in our analysis. (Additional Files 2 & 3).

## 3.5 Conclusion

Our SNP-based genetic linkage map of turkey with 531 markers reveals a genome length of 2,324 cM. This linkage map also allowed a comparison of the genome structures of turkey and chicken, demonstrating a very high degree of conservation in chromosome structure. A relatively low number of inter-, and intrachromosomal rearrangements was observed despite these two species being separated by 40 million years of evolution.

## Authors' contributions

MLA and MAMG analyzed the data. RPMAC organized the lab work and improved the paper with suggestions and comments. MLA wrote the paper and all other authors gave suggestions and comments for the improvement of paper. All authors read and approved the final manuscript.

## Acknowledgements

Breeding and Genomics Centre, Wageningen University) for their help on the lab work. We also thank Gerard Albers and the Breeding Research and Technology Centre at Hendrix Genetics, for their help in the design of the experiment as well as the collection of blood samples.

## Additional files:
### *Additional file 3.1*
**Link:** http://www.biomedcentral.com/1471-2164/11/647/additional
**Description:**
**Title: Detail of SBE primers along with their primer sequences and gene accession numbers.** This file contains PCR reverse and forward primer sequences along with the SNP specific SBE primer sequence. This file also contains gene name and their accession numbers.

### *Additional file 3.2*
**Link:** http://www.biomedcentral.com/1471-2164/11/647/additional
**Description:**
**Title: Linkage and physical maps (data) of turkey chromosomes along with the physical map of syntenic chicken chromosomes.** The detail of turkey linkage and physical maps along with the chicken physical map. This file also contains the flanking sequences of SNPs used in the present studied with their genotyping status.

### *Additional file 3.3*
**Link:** http://www.biomedcentral.com/1471-2164/11/647/additional
**Description:**
**Title: Linkage maps (Figures) of turkey chromosomes showing rearrangements with syntenic chicken chromosomes.** Figures showing comparative linkage maps of turkey and chicken including all the chromosomes mentioned in the present paper.

## References
1. **Food and agriculture organization statistical division (FAOSTAT) of the United Nations** [http://faostat.fao.org/]
2. Griffin D, Robertson L, Tempest H, Vignal A, Fillon V, Crooijmans R, Groenen M, Deryusheva S, Gaginskaya E, Carre W, Waddington D, Talbot R, Volker M, Masabanda J, Burt D: **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution**. *BMC Genomics* 2008, **9**(1):168.

3.    Reed KM, Chaves LD, Hall MK, Knutson TP, Harry DE: **A comparative genetic map of the turkey genome**. *Cytogenet Genome Res* 2005, **111**:118-127.

4.    Bumstead N, Palyga J: **A preliminary linkage map of the chicken genome**. *Genomics* 1992, **13**(3):690 - 697.

5.    Cheng HH, Levin I, Vallejo RL, Khatib H, Dodgson JB, Crittenden L, Hillel J: **Development of a genetic map of the chicken with markers of high utility**. *Poult Sci* 1995, **74**:1855 - 1874.

6.    Groenen MAM, Crooijmans RPMA, Veenendaal A, Cheng HH, Siwek M, Van der Poel JJ: **A comprehensive microsatellite linkage map of the chicken genome**. *Genomics* 1998, **49**:265-274.

7.    Herbergs J, Siwek M, Crooijmans RPMA, Poel JJvd, Groenen MAM: **Multicolour fluorescent detection and mapping of AFLP markers in chicken (Gallus domesticus)**. *Anim Genet* 1999, **30**:274-285.

8.    Groenen MAM, Cheng HH, Bumstead N, Benkel BF, Briles WE, Burke T, Burt DW, Crittenden LB, Dodgson J, Hillel J Lamont S, De Leon AP, Soller M, Takahashi H, Vignal A: **A consensus linkage map of the chicken genome**. *Genome Res* 2000, **10**:137 - 147.

9.    Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens H-J, Crooijmans RPMA, Lathrop FBM, Muir WM, Ka-Shu Wong G, Gut I, Andersson L: **A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate**. *Genome Res* 2009, **19**(3):510-519.

10.   Elferink M, van As P, Veenendaal T, Crooijmans R, Groenen M: **Regional differences in recombination hotspots between two chicken populations**. *BMC Genetics* 2010, **11**(1):11.

11.   Griffin DK, Robertson LBW, Tempest HG, Skinner BM: **The evolution of the avian genome as revealed by molecular cytogenetics**. *Cytogenet Genome Res* 2007, **117**:64-77.

12.   Stock AD, Bunch TD: **The evolutionary implications of chromosome-banding pattern homologies in the bird order Galliformes**. *Cytogenet Cell Genet* 1982, **34**:136-148.

13.   Shibusawa M, Nishibori M, Nishida-Umehara C, Tsudzuki M, Masabanda J, Griffin DK, Matsuda Y: **Karyotypic evolution in the Galliformes: An examination of the process of karyotypic evolution by comparison of the molecular cytogenetic findings with the molecular phylogeny**. *Cytogenetic and Genome Research* 2004, **106**(1):111-119.

70

14.     Ellegren H: **Evolutionary stasis: the stable chromosomes of birds**. *Trends in Ecology & Evolution* 2010, **25**(5):283-291.

15.     Nie W, O'Brien PCM, Ng BL, Fu B, Volobouev V, Carter NP, Ferguson-Smith MA, Yang F: **Avian comparative genomics: reciprocal chromosome painting between domestic chicken (Gallus gallus) and the stone curlew (Burhinus oedicnemus, Charadriiformes)—An atypical species with low diploid number**. *Chromosome Res* 2009, **17**(1):99-113.

16.     Guttenbach M, Nanda I, Feichtinger W, Masabanda JS, Griffin DK, Schmid M: **Comparative chromosome painting of chicken autosomal paints 1-9 in nine different bird species.** *Cytogenetic and Genome Research* 2003, **103**(1-2):173-184.

17.     Derjusheva S, Kurganova A, Habermann F, Gaginskaya E: **High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds**. *Chromosome Res* 2004, **12**:715-723.

18.     Heil J, Glanowski S, Scott J, Winn-Deen E, McMullen I, Wu L, Gire C, Sprague A: **An automated computer system to support ultra high throughput SNP genotyping**. In: *Pacific symposium on biocomputing: 2002*; 2002: 7:30-40.

19.     Kerstens HHD, Crooijmans RPMA, Veenendaal A, Dibbits BW, Chin-A-Woeng TFC, Dunnen JTd, Groenen MAM: **Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey**. *BMC Genomics* 2009, **10**:479.

20.     Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MAM, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J *et al*: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution**. *Nature* 2004, **432**:695 - 716.

21.     Zhang Z, Schwartz S, Wagner L, Miller W: **A greedy algorithm for aligning DNA sequences**. *J Comput Biol* 2000, **7**:203-214.

22.     Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W: **Human-mouse alignments with BLASTZ**. *Genome Res* 2003, **13**:103-107.

23.     Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova

M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS: **Highly parallel SNP genotyping**. *Cold Spring Harb Symp Quant Biol* 2003, **68**:69- 78.

24.    Hyten DL, Song Q, Choi I-Y, Yoon M-S, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB: **High-throughput genotyping with the GoldenGate assay in the complex genome of soybean**. *Theor Appl Genet* 2008, **116**:945-952.

25.    Green P, Falls K, Crooks S: **CRI-MAP Program**. In*.*, 2.4 edn: St. Louis: Washington university school of medicine; 1990.

26.    Stapley J, Birkhead TR, Burke T, Slate J: **A linkage map of the Zebra Finch Taeniopygia Guttata provides new insights into avian genome evolution**. *Genetics* 2008, **179**:651-667.

27.    Voorrips RE: **MapChart: Software for the graphical presentation of linkage maps and QTLs**. *J Hered* 2002, **93**(1):77-78.

28.    Korf I, Yandell M, Bedell J: **BLAST**. In*.*: O'Reilly Media 2003.

29.    Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Ann Blomberg L, Bouffard P, Burt DW, Crasta O, Crooijmans RPMA, Cooper K, Coulombe RA, De Supriyo, Delany ME, Dodgson JB, Dong JJ, Evans C, Frederickson KM, Flicek P, Florea L, Folkerts O, Groenen MAM, Harkins TT, Herrero J, Hoffmann S, Megens H-Jan, Jiang A, de Jong P, Kaiser P, Kim H *et al*: **Multi-Platform Next-Generation Sequencing of the Domestic Turkey (*Meleagris gallopavo*): Genome Assembly and Analysis**. *PLoS Biol* 2010, **8**(9):e1000475.

30.    Völker M, Backström N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, Griffin DK: **Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution**. *Genome Research* 2010, **20**(4):503-511.

31.    Kayang B, Fillon V, Inoue-Murayama M, Miwa M, Leroux S, Feve K, Monvoisin J-L, Pitel F, Vignoles M, Mouilhayrat C, Beaumont C, Ito S, Minvielle F. Vignal A: **Integrated maps in quail (Coturnix japonica) confirm the high degree of synteny conservation with chicken (Gallus gallus) despite 35 million years of divergence**. *BMC Genomics* 2006, **7**(1):101.

32.    Skinner BM, Robertson LBW, Tempest HG, Langley EJ, Ioannou D, Fowler KE, Crooijmans RPMA, Hall AD, Griffin DK, Völker M: **Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis**. *BMC Genomics* 2009, **10**:357.

33.     Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D, Sazanov A, Fries R, Waddington D: **The dynamics of chromosome evolution in birds and mammals**. *Nature* 1999, **402**(6760):411-413.

34.     Burt DW: **Origin and evolution of avian microchromosomes**. *Cytogenetic and Genome Research* 2002, **96**(1-4):97-112.

35.     Haldane JBS: **Sex-ratio and unisexual sterility in hybrid animals**. *J Genet* 1922, **12**:101-109.

36.     Huxley JS: **Sexual difference of linkage in Gammarus chereuxi**. *J Genet* 1928, **20**:145-156.

37.     Rodionov AV: **Micro vs. macro: Structural-functional organization of avian micro- and macrochromosomes**. *Genetika* 1996, **32**:597-608.

38.     Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D, Morrice DR, Law AS, Masabanda J, Sazanov A, Waddington D, Fries R, Burt DW: **Differences in gene density on chicken macrochromosomes and microchromosomes**. *Animal Genetics* 2000, **31**(2):96-103.

39.     Myers S, Bottolo L, Freeman C, McVean G, Donnelly P: **A fine-scale map of recombination rates and hotspots across the human genome**. *Science* 2005, **310**:321-324.

40.     Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J: **A High-resolution single nucleotide polymorphism genetic map of the mouse genome**. *PLoS Biol* 2006, **4**(12):e395.

41.     Versteeg R, van Schaik BD, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AH: **The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes.** . *Genome Res* 2003, **13**(9):1998-2004.

42.     Spencer CCA, Deloukas P, Hunt S, Mullikin J, Myers S, Silverman B, Donnelly P, Bentley D, McVean G: **The Influence of recombination on human genetic diversity**. *PLoS Genet* 2006, **2**(9):e148.

# 4

# Whole genome QTL mapping for growth, meat quality and breast meat yield traits in turkey

Muhammad L Aslam[1], John WM Bastiaansen[1], Richard PMA Crooijmans[1], Addie Vereijken[2], Martien AM Groenen[1]

[1] Animal Breeding and Genomics Centre, Wageningen University,6709PG, Wageningen, The Netherlands
[2] Hendrix Genetics, Research & Technology Centre, 5830 AC, Boxmeer, The Netherlands

## Abstract

**Background**

The turkey (*Meleagris gallopavo*) is an important agricultural species and is the second largest contributor to the world's poultry meat production. Demand of turkey meat is increasing very rapidly. Genetic markers linked to genes affecting quantitative traits can increase the selection response of animal breeding programs. The use of these molecular markers for the identification of quantitative trait loci, and subsequently fine-mapping of quantitative trait loci regions, allows for pinpointing of genes that underlie such economically important traits.

**Results**

The quantitative trait loci analyses of the growth curve, body weight, breast yield and the meat quality traits showed putative quantitative trait loci on 21 of the 27 turkey chromosomes covered by the linkage map. Forty-five quantitative trait loci were detected across all traits and these were found in 29 different regions on 21 chromosomes. Out of the 45 quantitative trait loci, twelve showed significant (p < 0.01) evidence of linkage while the remaining 33 showed suggestive evidence (p < 0.05) of linkage with different growth, growth curve, meat quality and breast yield traits.

**Conclusion**

A large number of quantitative trait loci were detected across the turkey genome, which affected growth, breast yield and meat quality traits. Pleiotropic effects or close linkages between quantitative trait loci were suggested for several of the chromosomal regions. The comparative analysis regarding the location of quantitative trait loci on different turkey, and on the syntenic chicken chromosomes, along with their phenotypic associations, revealed signs of functional conservation between these species.

## 4.1 Introduction

The turkey (*Meleagris gallopavo*, MGA) is an important agricultural species  and is the second largest contributor to the world's poultry meat production. Turkey stocks increased from 178 to 482 million and production volume increased from 1.2 to 5.6 M. tons between 1970 to 2008 [1]. This rapidly increasing demand of turkey meat motivated breeders and farmers to produce rapidly growing birds with a high market body weight (BW) and a desirable body conformation in order to maximize production efficiency and optimize production of preferred body cuts; e.g., breast muscle yield [2].

Commonly applied breeding programs for meat type birds, select for body weight (BW) and body composition traits (breast muscle yield, etc.), while  minimizing production costs. Recently, breeders have started measuring meat quality traits (drip loss, pH and color) as well as survival traits, at least in research project settings [3, 4]. Selection efforts have improved BW and body composition (i.e. increasing breast yield and lowering carcass fatness). These improvements, however, have also led to indirect and sometimes deleterious effects on meat quality and fitness traits [3]. Genetic parameters (heritabilities, genetic and phenotypic correlations) for the growth, meat quality and breast yield traits in turkey birds have been estimated [5], and showed unfavorable correlations of meat quality traits with the growth and the breast yield traits. The use of molecular markers that are directly or indirectly linked to QTL could provide potent tools to overcome these challenging correlations [6, 7]. In addition, identification and subsequent fine-mapping of QTL regions should allow for the pinpointing of genes that underlie such traits.

Several studies have indicated that knowledge about genetic markers linked to genes affecting quantitative traits can increase the selection response of animal breeding programs, especially for traits that are difficult to improve by traditional selection [8, 9]. Significant association between individual genetic markers and quantitative traits of economic importance have been reported in chicken [10-13] but no such reports exist for turkey.

A large number of studies are available on QTL mapping for the growth, meat quality and the body composition traits of chicken [7, 11, 14-16] showing significant effects of QTLs on these traits of economic importance in poultry breeding.

The detection of QTL and exploration of the underlying genes controlling these traits will benefit poultry breeding programs [17]. With this study we aim to build the same potential for turkey breeding programs by detecting quantitative trait loci for growth, meat quality and breast yield traits in turkey.

## 4.2 Methodology

### Resource population

Parents were randomly selected from two different commercial lines of turkey to produce F1 offspring [18]. Ten parent males were randomly selected from a high growth male line that contributed to a "large white product". Ten parent females were randomly selected from a high reproduction female line that contributed to a "heavy medium product". Average BW of males in the high growth line was 11.5 Kg and the average body weight of males in high reproduction line was 7.4 Kg at 14 weeks of age. Average egg production in the high growth line was 59.3 hatching eggs/24 weeks while average egg production in the high reproduction line was 115.5 hatching eggs/24 weeks. Parents were crossed to produce 10 full-sib families in the F1 generation. An F2 generation of 18 full sib families was produced by crossing 17 randomly selected F1 males and 18 randomly selected F1 females. One of the males was mated with two females; other F1 parents were mated only once. The F2 individuals were from 14 different hatches. In total, 973 F2 offspring were produced with an average full sib family size of 54.1 and a range of 31 to 90 individuals per family.

### Traits

Phenotypic data were recorded within a commercial breeding program. Body weight (BW), breast yield (BrY) and meat quality (MQ) traits were recorded on individuals of the F2 generation. Body weights were recorded at 1, 17, 40, 60, 80, and 120 days (BW01, BW17, BW40, BW60, BW80, and BW120, respectively). The breast meat yield traits; breast length (BrL), breast width (BrW), percent breast meat (PBM, *Pectoralis* (*P*) major and *P*. minor) and meat quality traits; percent drip loss (PDL), ultimate pH (pHu) and breast meat color (CIE L*a*b* system, where L* represents lightness, a* redness and b* yellowness) were measured at 20 weeks of age. These traits were measured as described previously [5].

Body weight observations at different time points were used to derive logistic growth curve traits i.e. asymptotic weight ($As_{wt}$), inflection point at which 50% of the asymptotic weight is achieved ($t_{mid}$), and a constant that is proportional to the overall growth rate (scale). The procedures and methodology for the estimation of these traits have previously been described [5].

### Genotype data and linkage map

The marker data and the linkage map utilized in the study were described in Aslam et al. [18]. The genotype data of 522 SNP, mapped to 27 turkey autosomes, was available after removal of uninformative and problematic SNP from the total set of 775 SNP [18]. The sex average linkage map was used, which had a length of 2164.8 cM with an average marker spacing of 4.4 cM. The data also included SNP that

were specifically selected from 5 different turkey genes; *PIT1*, *AFABP*, *PRKAG3*, *IGF2* and *GDF8.*

### *Statistical analysis*

***Descriptive analysis:*** Basic descriptive statistics, including number of observations (N), minimum values, maximum values, means and standard deviations (s.d.) were calculated by PROC MEANS of SAS software [19]. Fixed effects of sex and hatch were tested for significance on each trait with PROC GLM [19]. Effects that were found to be significant (P < 0.05) were included in the model for the QTL mapping analysis.

***QTL mapping:*** A regression-interval mapping method was applied which is available through the web-based software QTL EXPRESS accessed via the GridQTL portlet [20]. GridQTL is a portlet environment (available at http://www.gridqtl.org.uk/) that permits the analysis of computationally intensive datasets. Because of the full-sib structure in the F2, and the absence of genotypes on the parent generation, the analyses were carried out by applying a sib-pair model. Sex and hatch (n = 14) effects were tested for all traits and included in the model only if statistically significant (P < 0.05).

F-statistic profiles were generated at 1 cM intervals along each chromosome to identify the most likely QTL position. Significance thresholds were determined by permutation of the dataset [21], with 10,000 permutations performed to obtain single position as well as chromosome-wide significance levels. QTL that exceeded the chromosome-wide F-critical threshold at a P < 0.05 were reported as suggestive QTL, while exceeding a chromosome-wide F-critical threshold of P < 0.01 was considered evidence for a significant QTL effect. QTL variance estimates were obtained from a separate regression analysis of squared differences on IBD sharing of full-sibs at the QTL positions [22].

On each chromosome, regions were defined based on the occurrence of QTL. Two or more QTL were considered to be located in the same region if the distance between the chromosomal positions of these QTL was equal or less than 10 cM.

***Comparative QTL mapping:*** All significant as well as all suggestive turkey QTL were mirrored on the chicken genome. Nucleotide positions of SNP flanking the turkey QTL were mapped to chicken chromosomes and the chicken nucleotide positions were subsequently used to obtain cM positions on the chicken genome [18] that correspond to the positions of QTL discovered in turkey. These chicken genome positions of turkey QTL were compared to chicken QTL positions for the same trait, or a very similar trait, which were obtained from QTLdb [23]. The distance of the turkey QTL position on the chicken map to the nearest chicken QTL for the same trait was calculated.

To test whether QTL are conserved between chicken and turkey we used the distance from a random chicken map position to a chicken QTL as our null hypothesis. Under the null hypothesis, chicken linkage map positions (cM) were randomly chosen (n = 100) and their average distance to BW QTL from the chicken QTLdb was calculated. The distance between randomly selected positions from the chicken linkage map and the nearest QTL position from QTLdb were averaged and compared to the average distance between chicken and turkey QTL for the same trait.

### *Ethical approval for the use of animals in this study*

Although animals were used in this experimental work, no direct experiments were performed on them. Blood sample collection was carried out by licensed and authorized personnel under approval of Hendrix Genetics. No approval from the ethics committee was necessary.

## 4.3 Results

### *Descriptive Analysis*

A descriptive analysis of all the traits under study is summarized in Table 4.1. The effect of sex was significant ($P < 0.0005$) for all the traits except for the weight of 1 day old chicks (BW01), percent breast meat (PBM) and the redness of meat (a*). The effect of hatch was also significant for all the traits.

### *QTL mapping*

QTL that surpassed the suggestive or significant linkage threshold were summarized in Tables 2, 3, 4 & 5. The QTL analyses for the growth curve (Table 4.2), BW (Table 4.3), BY (Table 4.4) and the MQ traits (Table 4.5) showed putative QTL on 21 of the 27 turkey chromosomes covered by the linkage map. Forty-five QTL were detected across all traits and these were found in 29 different regions on 21 chromosomes. Out of the 45 QTL, twelve QTL showed significant ($p < 0.01$) evidence of linkage while the remaining 33 QTL showed suggestive evidence ($p < 0.05$) of linkage with different growth, growth curve, meat quality and breast yield traits.

MGA3 appeared to be important for all trait groups except BW traits, with four different regions affecting $As_{wt}$, BrW, b* and PDL at 92, 132, 107 and 65 cM respectively (Table 4.2, 4.4 & 4.5). The QTL for b* on chromosome 3 was found significant, the others were suggestive. The four QTL affected four different traits and their positions were also in different regions which suggests that four different QTL were involved, one for each of the traits.

**Table 4.1:** Descriptive statistics, including the estimates for the significant fixed effects (Sex and Hatch).

| Traits(units) | N | Minimum | Maximum | LS Mean | RSD | Sex[1] | Hatch[2] |
|---|---|---|---|---|---|---|---|
| BW01(Kg) | 810 | 0.04 | 0.07 | 0.06 | 0.06 | 0.00 | 0.02* |
| BW17(Kg) | 785 | 0.08 | 0.60 | 0.33 | 0.43 | 0.22* | 0.13* |
| BW40(Kg) | 751 | 0.52 | 2.32 | 1.35 | 1.64 | 0.69* | 0.31* |
| BW60(Kg) | 710 | 1.50 | 4.96 | 3.11 | 3.65 | 1.27* | 0.55* |
| BW80(Kg) | 693 | 3.06 | 8.50 | 5.45 | 6.33 | 2.25* | 1.19* |
| BW120(Kg) | 655 | 4.54 | 15.90 | 10.39 | 12.19 | 5.04* | 1.50* |
| PBM (%) | 785 | 0.02 | 13.40 | 10.73 | 2.15 | 0.10 | 8.83* |
| BrL(mm) | 937 | 155.00 | 300.00 | 212.57 | 28.53 | 48.30* | 21.29* |
| BrW(mm) | 937 | 109.00 | 203.00 | 146.88 | 16.17 | 25.68* | 21.39* |
| PDL (%) | 828 | 2.21 | 14.10 | 5.09 | 1.28 | 0.94* | 1.35* |
| pHu | 838 | 5.26 | 6.02 | 5.75 | 0.11 | 0.04* | 0.53* |
| L* | 864 | 40.30 | 53.60 | 45.92 | 1.82 | 0.98* | 2.65* |
| a* | 864 | 1.30 | 9.20 | 5.27 | 1.00 | 0.09 | 2.56* |
| b* | 864 | 0.10 | 5.60 | 2.28 | 0.84 | 0.54* | 0.63* |
| $As_{wt}$(Kg) | 645 | 4.65 | 20.23 | 12.29 | 3.47 | 6.50* | 2.92* |
| $T_{mid}$(Day) | 645 | 59.86 | 112.24 | 82.85 | 5.44 | 6.14* | 11.75* |
| Scale(Day) | 645 | 12.66 | 29.15 | 20.61 | 2.03 | 1.95* | 5.13* |

N = Number of records; minimum = minimum values; maximum = maximum values; LS Mean = least square mean; RSD = residual standard deviation; BW01, BW17, BW40, BW60, BW80 ,and BW120 are the BW at days 1,17, 40, 60, 80, and 120 of age, respectively; PBM = percentage breast meat at 20 week of age; BrL = breast length at 20 week of age; BrW = breast width at 20 wk of age; PDL = percent drip loss at 20 week of age; pHu = ultimate pH at 20 wk of age; L* = lightness at 20 wk of age; a* = redness at 20 wk of age; b* = yellowness at 20 wk of age; $As_{wt}$ = upper asymptotic weight (estimated growth curve parameter); $t_{mid}$ = inflection point at 50% asymptote (estimated growth curve parameter); scale = constant that is proportional to the overall growth rate (estimated growth curve parameter). [1] = Difference between sexes in the Least square means (LS Means) of the traits. [2] = Difference between the maximum and minimum LS Means of the traits with respect to the week of hatch.
*$P \le 0.0005$

Two QTL regions were detected on chromosome 5, the first region showed a QTL for development in weight (BW17 and BW40) at 60-63 cM, and the second region showed a QTL for BrL at 113 cM. The QTL for BrL was in a separate region. Another region with QTL for development in BW traits (BW40, BW60 and BW80) was located on chromosome 8 at cM position 1 (Figure 4.1 & 4.2).

Two regions on MGA12, the first with QTL affecting weight development (BW40 and BW80) and the second with QTL affecting the quality of meat (b*, and PDL) were detected at 0 to 1 and 17 to 27 cM respectively (Table 4.3 & 4.5).

In our study, MGA22 showed multiple QTL affecting growth (growth curve and BW traits) as well as a QTL with an effect on PBM. A QTL at position 0 to 6 cM showed significant evidence ($p < 0.01$) for an effect on the growth curve trait scale, while at the same position suggestive evidence ($p < 0.05$) was found for an effect on the other growth traits BW40, BW120, $As_{wt}$, and $t_{mid}$ as well as an effect on PBM (Table 4.2, 4.3 & 4.4).

Again, multiple QTL were detected on chromosome 28 with significant effects on $As_{wt}$, BW120 and BrL and with suggestive evidence for BrW (Table 4.2, 4.3 & 4.4) with QTL positions between 0 and 12 cM.

When focusing on meat quality, QTL with significant effects ($p < 0.01$) on meat quality, yellowness (b*), were detected on chromosome 3, 12 and 26 at position 107 cM, 27 cM and 43 cM respectively (Table 4.5). Additional QTL with suggestive effects on percent drip loss were detected on chromosome 1, 3 and 12 at position 71, 65 cM and 17 cM respectively (Table 4.5). Suggestive evidence of a QTL affecting lightness (L*) of meat was also detected on chromosome 8 at cM position 1 (Figure 4.2). No significant QTL was detected for redness (a*) and the ultimate pH (pHu) of meat (Additional file 4.1 & 4.2).

### *Comparative QTL mapping*

For seven out of the 15 turkey QTL that affected BW traits, QTL were found for the same or a very similar trait on syntenic regions in the chicken genome, within a distance of 8 cM or less. The average distance between syntenic positions of the turkey BW QTL in chicken and the nearest chicken QTL positions (from QTLdb) was 14.7 cM (Additional file 4.3). The seven turkey QTL with nearby syntenic chicken QTL were detected on MGA1, 5, 13, 20 and MGA22. A turkey QTL affecting b* was also found nearby a chicken QTL for b* with a distance of less than 7 cM between the syntenic QTL positions in these species. This QTL for b* was detected on MGA12 (Additional file 4.3).

The distance from a randomly selected positions (n = 100) on the chicken linkage map to the nearest chicken QTL for BW traits was on average 18.06 ± 3.08 cM (Additional file 4.3).

**Table 4.2:** QTL mapped on different chromosomes of turkey affecting growth curve traits.

| Trait | Chromosome | Location (cM) | qtlV | F-Statistics | Flanking Markers | F-Statistics Threshold[1] | |
|---|---|---|---|---|---|---|---|
| | | | | | | P<0.05 | P<0.01 |
| Scale | MGA2 | 113 | 0.07 | 15.40 | B002042-A004960 | 10.03 | 16.35 |
| $As_{wt}$ | MGA3 | 92 | 0.09 | 11.91 | A005884-A001055 | 10.53 | 17.82 |
| $As_{wt}$ | MGA13 | 49 | 0.09 | 12.16* | A002976-B002771 | 6.67 | 11.04 |
| Scale | MGA15 | 30 | -0.07 | 14.59* | B002847-A003255 | 8.89 | 13.52 |
| Aswt | MGA22 | 2 | 0.10 | 12.30 | A000901-A006033 | 6.58 | 12.83 |
| $T_{mid}$ | MGA22 | 6 | -0.02 | 7.61 | A003266-A000012 | 6.80 | 13.42 |
| Scale | MGA22 | 5 | 0.05 | 10.91* | A006033-A003266 | 6.46 | 10.56 |
| $As_{wt}$ | MGA28 | 16 | 0.17 | 18.70* | B000023-B001881 | 5.05 | 8.12 |

* QTLs with significant evidence ($P < 0.01$). [1]Chromosome wide significance thresholds from permutation test.

**Table 4.3:** QTL mapped on different chromosomes of turkey affecting body weight traits.

| Trait | Chromosome | Location (cM) | qtlV | F-Statistics | Flanking Markers | F-Statistics Threshold[1] | |
|---|---|---|---|---|---|---|---|
| | | | | | | P<0.05 | P<0.01 |
| BW40 | MGA1 | 217 | 0.03 | 11.05 | B003270-A005799 | 10.81 | 15.82 |
| BW17 | MGA5 | 63 | 0.15 | 10.24 | A001354-A005103 | 8.89 | 16.02 |
| BW40 | MGA5 | 60 | 0.11 | 10.40 | A001354-A005103 | 9.22 | 15.26 |
| BW40 | MGA8 | 1 | 0.11 | 11.49 | B000608-A001480 | 7.29 | 11.63 |
| BW60 | MGA8 | 1 | 0.06 | 9.05 | B000608-A001480 | 7.17 | 11.51 |
| BW80 | MGA8 | 1 | 0.07 | 11.95* | B000608-A001480 | 7.29 | 11.62 |
| BW40 | MGA12 | 0 | 0.18 | 8.39 | B000094-B000257 | 5.91 | 10.40 |
| BW80 | MGA12 | 1 | 0.13 | 8.47 | B000094-B000257 | 6.26 | 10.98 |
| BW120 | MGA13 | 54 | 0.05 | 7.81 | A002976-B002771 | 6.95 | 12.86 |
| BW40 | MGA20 | 51 | 0.04 | 7.93 | B002015-B002517 | 6.73 | 11.78 |
| BW120 | MGA22 | 0 | 0.12 | 12.15 | B002897-A000901 | 7.52 | 13.29 |
| BW40 | MGA22 | 6 | 0.08 | 8.78 | A003266-A000012 | 7.60 | 12.39 |
| BW40 | MGA26 | 0 | 0.15 | 9.54 | B000407-B002784 | 8.23 | 16.83 |
| BW120 | MGA28 | 12 | 0.11 | 11.04* | B000023-B001881 | 5.80 | 10.23 |
| BW01 | MGA30 | 0 | -0.01 | 9.08* | B003031-B000504 | 4.39 | 9.07 |

* QTLs with significant evidence ($P < 0.01$). [1]Chromosome wide significance thresholds from permutation test.

**Table 4.4:** QTL mapped on different chromosomes of turkey affecting breast yield traits.

| Trait | Chromosome | Location (cM) | qtlV | F-Statistics | Flanking Markers | F-Statistics Threshold[1] | |
|---|---|---|---|---|---|---|---|
| | | | | | | P<0.05 | P<0.01 |
| BrW | MGA3 | 132 | 0.15 | 16.64 | B003202-B002875 | 10.87 | 16.67 |
| PBM | MGA4 | 29 | 0.18 | 10.88 | A006113- B001871 | 8.27 | 11.94 |
| BrL | MGA5 | 113 | 0.06 | 9.19 | A003231-A000813 | 8.57 | 14.63 |
| PBM | MGA11 | 36 | 0.30 | 9.65 | B002433-A003945 | 7.39 | 12.49 |
| PBM | MGA19 | 41 | 0.14 | 7.31[*] | B002491-B002546 | 3.41 | 5.95 |
| PBM | MGA22 | 6 | 0.22 | 10.49 | A003266-A000012 | 7.78 | 12.24 |
| PBM | MGA26 | 45 | 0.38 | 9.25 | B002264-A006279 | 8.20 | 14.15 |
| BrL | MGA28 | 4 | -0.01 | 17.10[*] | B000278-B000023 | 5.27 | 9.32 |
| BrW | MGA28 | 0 | -0.01 | 5.86 | B000278-B000023 | 5.13 | 8.26 |

[*] QTLs with significant evidence (P < 0.01). [1]Chromosome wide significance thresholds from permutation test.

**Table 4.5:** QTL mapped on different chromosomes of turkey affecting meat quality traits.

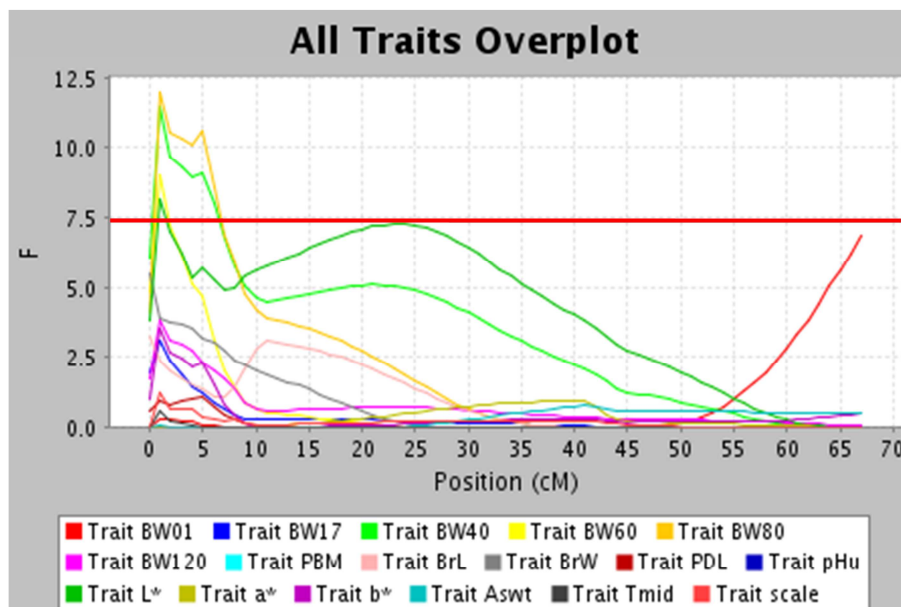| Trait | Chromosome | Location (cM) | qtlV | F-Statistics | Flanking Markers | F-Statistics Threshold[1] | |
|---|---|---|---|---|---|---|---|
| | | | | | | P<0.05 | P<0.01 |
| PDL | MGA1 | 71 | 0.17 | 15.57 | B001935-B001936 | 10.44 | 15.90 |
| b* | MGA3 | 107 | 0.07 | 16.66[*] | A002870-B003116 | 9.41 | 15.02 |
| PDL | MGA3 | 65 | 0.11 | 10.61 | B003023-B002640 | 8.91 | 12.62 |
| b* | MGA4 | 30 | 0.1 | 9.88 | B001871-B002284 | 8.72 | 13.46 |
| PDL | MGA7 | 0 | 0.1 | 10.62 | A001382-B002403 | 7.98 | 13.70 |
| L* | MGA8 | 1 | 0.07 | 8.16 | B000608-A001480 | 7.37 | 12.12 |
| b* | MGA12 | 27 | 0.08 | 28.46[*] | A004841-A004198 | 6.44 | 9.87 |
| PDL | MGA12 | 17 | 0.06 | 5.99 | A001153-B000396 | 5.37 | 8.36 |
| PDL | MGA14 | 55 | 0.08 | 9.39 | A003474-B002743 | 7.38 | 12.29 |
| PDL | MGA17 | 52 | 0.19 | 10.92 | A003133-A000203 | 7.54 | 13.18 |
| b* | MGA21 | 61 | 0.08 | 11.53 | B003125-A004009 | 8.34 | 15.02 |
| PDL | MGA24 | 30 | 0.06 | 7.11 | B000536-B002896 | 5.41 | 10.38 |
| b* | MGA26 | 43 | 0.1 | 17.06[*] | B002430-B002264 | 8.09 | 15.25 |

[*] QTLs with significant evidence (P < 0.01). [1]Chromosome wide significance thresholds from permutation test.

**MGA5**
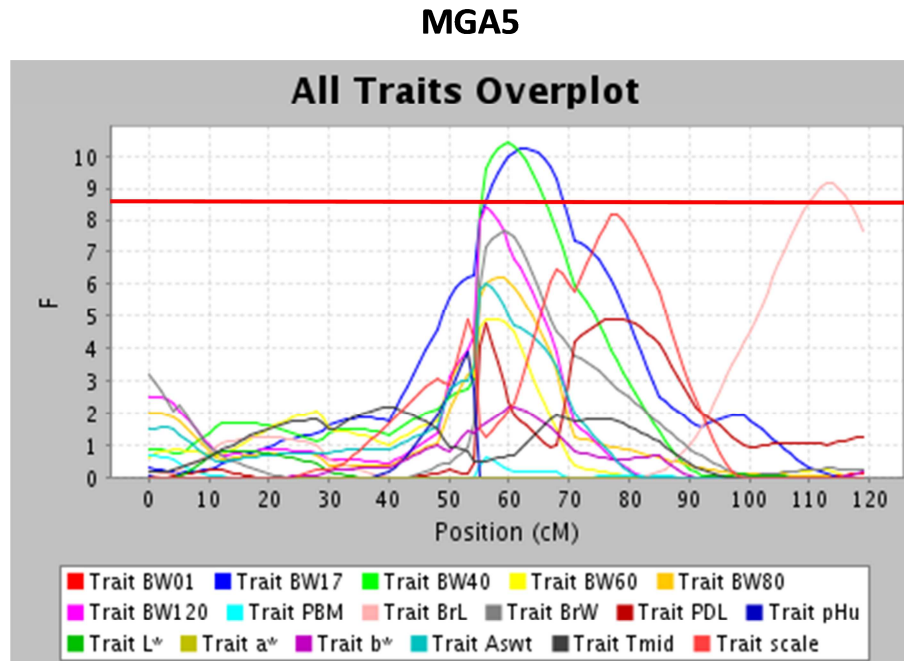


‒ F-statistical threshold (P<0.05) for BrL.

**Figure 4.1.** Identified QTL on turkey chromosome 5 affecting growth, meat quality and breast yield traits. BW01, BW17, BW40, BW60, BW80, and BW120 are the BW at days 1,17, 40, 60, 80, and 120 of age; PBM = percentage breast meat at 20 wk of age; BrL = breast length at 20 wk of age; BrW = breast width at 20 wk of age; PDL = percent drip loss at 20 wk of age; pHu = ultimate pH at 20 wk of age; L* = lightness at 20 wk of age; a* = redness at 20 wk of age; b* = yellowness at 20 wk of age; $As_{wt}$ = upper asymptote (estimated growth curve parameter); $t_{mid}$ = inflection point at 50% asymptote (estimated growth curve parameter); scale = constant that is proportional to the overall growth rate (estimated growth curve parameter).

## MGA8



- F-statistics threshold (P<0.05) for BW60

**Figure 4.2.** Identified QTL on turkey chromosome 8 affecting growth, meat quality and breast yield traits. BW01, BW17, BW40, BW60, BW80, and BW120 are the BW at days 1,17, 40, 60, 80, and 120 of age; PBM = percentage breast meat at 20 wk of age; BrL = breast length at 20 wk of age; BrW = breast width at 20 wk of age; PDL = percent drip loss at 20 wk of age; pHu = ultimate pH at 20 wk of age; L* = lightness at 20 wk of age; a* = redness at 20 wk of age; b* = yellowness at 20 wk of age; $As_{wt}$ = upper asymptote (estimated growth curve parameter); $t_{mid}$ = inflection point at 50% asymptote (estimated growth curve parameter); scale = constant that is proportional to the overall growth rate (estimated growth curve parameter).

## 4.4 Discussion

QTL were detected for growth, breast yield and meat quality traits which are important traits in poultry breeding. This study adds important new information from a genome wide search for QTL in turkeys, and is the first to report the detection and positioning of loci affecting commercially important traits in turkeys. Several chromosomes showed multiple QTL at nearby positions, indicating that pleiotropic effects may be playing a role. We expected to find overlapping QTL positions for multiple BW traits because these traits were previous found to have high genetic correlations among each other [5]. In the present study, eight QTL

were detected with a significant effect on growth. For seven of these eight QTL, additional significant or suggestive QTL for other growth traits were detected in the same chromosome region. This presence of multiple QTL for genetically correlated traits suggests the presence of QTL with pleiotropic effects on these traits. A good example is the identification of QTL for $As_{wt}$ and BW120 in the same region of chromosomes 13, 22 and 28. Traits $As_{wt}$ and BW120 are very similar traits that both represent mature BW and have a high genetic correlation of nearly 1 [5].

Comparative studies of turkey and chicken based on cytogenetic [24], genome sequence [25], and linkage [18] analyses have shown highly conserved karyotypes and genomic structure between these species. In the present study, a number of traits were found to be affected by QTL on MGA22 including BW traits. MGA22 appeared to play a role in the genetic variation of growth patterns in turkey, harboring a QTL with an effect on all three growth curve traits ($As_{wt}$, $t_{mid}$ and scale). QTL models were fitted on growth curve parameters to estimate effects on parameters that can be interpreted for their biologically meaning in the growth pattern, in addition to results from applying QTL models on BW observations at different time points. Applying QTL models on BW observations estimates the effect of a QTL on weight at that particular age while applying QTL model on growth curve parameters may give insight in the effect a of QTL throughout the growth pattern of an individual [26]. The QTL affecting the BW traits on chromosome 22 of turkey are located at a position syntenic to a region on GGA20 which was previously shown to contain a QTL for growth [15, 27] (Additional file 4.3). Likewise, the region on MGA1 containing the QTL for PDL is syntenic to a region on GGA1 also shown to contain a QTL for the same trait [14].

The identification of QTLs affecting BW traits on MGA1, 5, 13, 20, MGA22 and a QTL affecting meat color trait (b*) on MGA12 are also in agreement with the QTL reported for these  traits on the syntenic GGA1, 5, 11, 18, 20, and GGA10 respectively [27-30]. A high level of structural genomic conservation has been identified between turkey and chicken [18, 24, 25]. The comparison of turkey QTL positions, mirrored on the chicken genome, with the chicken QTL positions for the same trait suggests that in addition to the structural genomic conservation, functional genomic conservation also exist between these species.

The SNPs that are located within growth related genes (*PIT1*, *AFABP*, *PRKAG3*, *IGF2* and *GDF8)* were used to test for direct effects of these SNPs on the growth traits. When these SNPs were included as fixed effects in the model, the F-value at the position of these SNPs decreased by more than 50%. The large impact of these SNPs on the QTL model does not necessarily mean that the SNPs are causative mutations, but these SNPs explain an important amount of QTL variation, either

directly or through LD with the causative mutations. The candidate genes (*PIT1*, *AFABP*, *PRKAG3*, *IGF2* and *GDF8)* were known to affect growth related traits in other species making it likely that these are the actual genes underlying the QTL effects, even though LD extends over large regions [31] and the other genes in the neighborhood cannot be excluded.

Estimates of QTL variance were not obtained from the QTLexpress analysis output. To estimate the variance explained by each QTL, the regression slopes were used to calculate QTL variances (qtlV) as a proportion of the residual variance. These estimates of QTL variance are likely to be overestimates [32], but for a few QTL a negative QTL variance estimate was obtained because the regression slopes were positive in the regressions used to estimate them.

To search for positional candidate genes near the QTL, the sequence annotation of turkey was used. The Positions (cM) of the SNPs flanking the significant turkey QTLs, as well as the sequence surrounding the SNPs, were used to convert the cM positions of QTL on the linkage map into base-pair (bp) positions on the turkey genome. First the sequences around SNPs that flank the turkey QTL were used to obtain the position (bp) of these SNPs in the turkey genome [18]. Subsequently, the approximate position (bp) of turkey QTL in the turkey genome was predicted by using the relative distances in cM of the turkey QTL to the flanking SNP positions. Then these same relative distances were applied to the interval between the turkey genome positions (bp) of the flanking SNPs. Finally, functional information was inspected for genes within a region of ± 500Kb from the predicted QTL positions (bp) for the 10 longest chromosomes and within ± 100Kb for the 20 smallest chromosomes. Near most QTL, genes were found with unknown function or functions related to metabolism or transcription and translation processes. These genes can be responsible for the QTL effects that were found but no conclusion can be drawn. No genes were found on MGA22 within the window of ± 100Kb from the QTL position (bp) (Additional file 4.4).

As described earlier, the turkey QTL positions (bp) were mirrored onto the chicken genome. Genes on the chicken genome were identified within the same window ranges as applied in turkey. Two potential candidate genes were found in chicken for turkey QTL, namely EYA1 and Col5A1 which have functions in morphogenesis (drosophila) [33, 34] and fibrillogenesis [35] respectively. The genes EYA1 and Col5A1 were present in the syntenic turkey chromosomes but were positioned at 1345 Kb and 300.4 Kb away from the QTL positions (bp) in the turkey genome which were outside of selected search window for candidate genes.

Potentially pleiotropic effects of QTLs were observed in a number of regions of different turkey chromosomes. A QTL for PBM was found on chromosome 22 near

the QTL for BW and the QTL for the growth curve traits which could probably be explained by a pleiotropic effect of this QTL. In our study, PBM was recorded as a single trait, combining *P.* major and *P.* minor instead of measuring *P.* major and *P.* minor as two separate traits as suggested by Ankra-Badu et al. [36] on chicken who suggested that *P.* major and *P.* minor should be treated separately because these traits were found to be influenced by different QTL [36].

QTL for the breast yield traits, BrL and BrW, were found co-located on chromosome 28 which also harbored QTL for growth traits BW120 and As$_{wt}$, all within a range of 16 cM. These results fit expectations that were based on the high genetic correlation among BrL and BrW with BW traits and As$_{wt}$ [5].

No significant QTL were detected for pHu and a*. Some regions on chromosomes 1, 4, 5, 16 (pHu) and 2, 3, 6 (a*) did show an effects on these two traits (pHu and a*) but the observed F-value for these region did not surpass the threshold (Additional files 4.1 & 4.2). Given the high genetic correlation between PDL and pHu [5], QTL for pHu may have been expected on at least a part of the same chromosomes where QTL for PDL were detected. This lack of concordance may indicate that partially different sets of genes are involved in the control of these traits and/or that there were differences in power to detect QTL for these traits.

A QTL for L* was found in the same region as QTL for BW traits on chromosome 8. Similar to the breast yield traits, L* also had high genetic correlation with BW traits [5] which can be interpreted as an indication towards a pleiotropic nature of this QTL on chromosome 8.

Quality of meat is of interest to breeders and the identification of QTLs, markers and genes associated with meat characteristics would be of great value to improve the meat quality traits which are shown to have reasonable heritabilities (0.09-0.30) in turkeys [5]. In the present study, significant QTL for meat color trait (b*) were detected on three different chromosomes (3, 12, and 26) and suggestive QTL on two additional chromosomes (4 and 21). QTL for PDL were also found on two of these chromosomes (3 and 12). The QTL for PDL on chromosome 3 is, however, located at a distance from the QTL for b* while on chromosome 12, the QTL for PDL was observed in the same region as the QTL for b*. These results are also in agreement with the high genetic correlation between b* and PDL [5].

## 4.5 Conclusion

A large number of QTL were detected across the turkey genome, which affected growth, breast yield and meat quality traits. Pleiotropic effects or close linkages between QTL were suggested for several of the chromosomal regions. The comparative analysis regarding the location of QTL on different turkey and the

syntenic chicken chromosomes, in combination with their association with phenotype revealed signs of functional conservation between these species.

## Authors' contributions

MLA analyzed the data. MLA and JWMB wrote the paper and all other authors gave suggestions and comments for the improvement of paper. All authors read and approved the final manuscript.

## Acknowledgements

## Additional files:

### *Additional file 4.1*

**Link:** http://www.biomedcentral.com/1471-2156/12/61/additional

**Description:**

**Title: QTL (data) regions affecting growth, breast yield and meat quality traits mapped on different turkey chromosomes.** Details of QTL regions from all turkey chromosomes with F-statistics using chromosome wide F-statistics threshold.

### *Additional file 4.2*

**Link:** http://www.biomedcentral.com/1471-2156/12/61/additional

**Description:**

**Title: QTL (figures) regions affecting growth, breast yield and meat quality traits mapped on different turkey chromosomes.** Peaks showing QTL on different turkey chromosomes with an effect on growth curve, breast yield, body weight and meat quality traits.

### *Additional file 4.3*

**Link:** http://www.biomedcentral.com/1471-2156/12/61/additional

**Description:**

**Title: Comparative QTL mapping between turkey and chicken and predicted underlying genes.** Data file with comparative data for QTL positions on different

turkey chromosomes with the projection of these positions on different chicken chromosomes.

### *Additional file 4.4*

**Link:** http://www.biomedcentral.com/1471-2156/12/61/additional

**Description:**

**Title: List of genes found within the selected window across the significant QTL positions.** This file contains the names of underlying genes within the range of ± 500 Kb (Ten largest chromosomes) and ± 100 Kb (remaining chromosomes) from the projected QTL positions at the different turkey and chicken chromosomes.

## References

1. **Food and agriculture organization statistical division (FAOSTAT) of the United Nations [http://faostat.fao.org/]**

2. Le Bihan-Duval E, Debut M, Berri CM, Sellier N, Santé-Lhoutellier V, Jégo Y, Beaumont C: **Chicken meat quality : genetic variability and relationship with growth and muscle characteristics.** *BMC Genetics* 2008, **9**:53.

3. Berri C, Wacrenier N, Millet N, Le Bihan-Duval E: **Effect of selection for improved body composition on muscle and meat characteristics of broilers from experimental and commercial lines.** *Poult Sci* 2001, **80**:833-838.

4. Li H, Deeb N, Zhou H, Mitchell AD, Ashwell CM, Lamont SJ: **Chicken quantitative trait loci for growth and body composition associated with transforming growth factor-β genes1.** *Poult Sci* 2003, **82**:347-356.

5. Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Ducro BJ, Vereijken A, Megens H-J, Groenen MAM: **Genetic variances, heritabilities and maternal effects on body weight, breast meat yield, meat quality traits and the shape of the growth curve in turkey birds.** *BMC Genetics* 2011, **12**(1):14.

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

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

8. Meuwissen THE, Goddard ME: **The use of marker-haplotypes in animal breeding schemes.** *Genet Sel Evol* 1996, **28**:161-176.

9.      Pyasatian N, Fernando RL, Dekkers JCM: **Genomic selection for marker-assisted improvement in line crosses.** *Theor Appl Genet* 2007, **115**:665-674.

10.     Yonash N, Bacon LD, Witter RL, Cheng HH: **High resolution mapping and identification of new quantitative trait loci (QTL) affecting susceptibility to Marek's disease.** *Anim Genet* 1999, **30**:126-135.

11.     van Kaam JBCHM, Groenen MAM, Bovenhuis H, Veenendaal A, Vereijken ALJ, van Arendonk JAM: **Whole genome scan in chickens for quantitative trait loci affecting growth and feed efficiency.** *Poult Sci* 1999a, **78**:15-23.

12.     van Kaam JBCMH, Groenen MAM, Bovenhuis H, Veenendaal A, Vereijken ALJ, van Arendonk JAM: **Whole genome scan in chickens for quantitative trait loci affecting carcass traits.** *Poult Sci* 1999b, **78**:1091-1099.

13.     Yonash N, Cheng HH, Hillel J, Heller DE, Cahaner A: **DNA microsatellites linked to quantitative trait loci affecting antibody response and survival rate in meat-type chickens.** *Poult Sci* 2001, **80**:22-28.

14.     Nadaf J, Gilbert H, Pitel F, Berri CM, Feve K, Beaumont C, Duclos MJ, Vignal A, Porter TE, Simon J *et al*: **Identification of QTL controlling meat quality traits in an F2 cross between two chicken lines selected for either low or high growth rate.** *BMC Genomics* 2007, **8**:155.

15.     Wahlberg P, Carlborg Ö, Foglio M, Tordoir X, Syvänen A-C, Lathrop M, Gut IG, Siegel PB, Andersson L: **Genetic analysis of an F2 intercross between two chicken lines divergently selected for body-weight.** *BMC Genomics* 2009, **10**:248.

16.     Ankra-Badu GA, Le Bihan-Duval E, Mignon-Grasteau S, Pitel F, Beaumont C, Duclos MJ, Simon J, Carre W, Porter TE, Vignal A *et al*: **Mapping QTL for growth and shank traits in chickens divergently selected for high or low body weight.** *Anim Genet* 2010, **41**:400-405.

17.     Ou JT, Tang SQ, Sun DX, Zhang Y: **Polymorphisms of three neuroendocrine-correlated genes associated with growth and reproductive traits in the chicken.** *Poult Sci* 2009, **88** 722-727.

18.     Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Vereijken A, Megens H-J, Groenen MAM: **A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes.** *BMC Genomics* 2010, **11**:647.

19.     SAS Institute: **SAS User's Guide: Statistics.** In.: SAS Institute Inc., Cary, NC; 2000.

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

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

22. Haseman JK, Elston RC: **The investigation of linkage between a quantitative trait and a marker locus.** *Behav Genet* 1972, **2**:3-19.

23. Hu Z-L, Reecy JM: **Animal QTLdb: Beyond a Repository - A Public Platform for QTL Comparisons and Integration with Diverse Types of Structural Genomic Information.** *Mammalian Genome* 2007, **18**:1-4.

24. Griffin D, Robertson L, Tempest H, Vignal A, Fillon V, Crooijmans R, Groenen M, Deryusheva S, Gaginskaya E, Carre W *et al*: **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution.** *BMC Genomics* 2008, **9**(1):168.

25. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Ann Blomberg L, Bouffard P, Burt DW, Crasta O, Crooijmans RPMA *et al*: **Multi-Platform Next-Generation Sequencing of the Domestic Turkey *Meleagris gallopavo*):  Genome Assembly and Analysis.** *PLoS Biol* 2010, **8**(9):e1000475.

26. Li Y, Wu R: **Functional mapping of growth and development.** *Biol Rev* 2010, **85**:207-216.

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

28. Carlborg O, Kerje S, Schütz K, Jacobsson L, Jensen P, Andersson L: **A global search reveals epistatic interaction between QTL for early growth in the chicken.** *Genome Res* 2003, **13**(3):413-421.

29. Wright D, Kerje S, Lundström K, Babol J, Schütz K, Jensen P, Andersson L: **Quantitative trait loci analysis of egg and meat production traits in a red junglefowl x White Leghorn cross.** *Anim Genet* 2006, **37**(6):529-534.

30. Zhou H, Deeb N, Evock-Clover CM, Ashwell CM, Lamont SJ: **Genome-wide linkage analysis to identify chromosomal regions affecting phenotypic traits in the chicken. I. Growth and average daily gain.** *Poult Sci* 2006, **85**(10):1700-1711.

31. Kruglyak L: **What is significant in whole-genome linkage disequilibrium studies?** *Am J Hum Genet* 1997(61):810-812.

32.     Beavis WD: **QTL analyses: power, precision, and accuracy. In:** *Molecular dissection of complex traits.* Edited by Paterson AH. New York: CRC Press; 1998: 145-162.

33.     Liu YH, Jakobsen JS, Valentin G, Amarantos I, Gilmour DT, Furlong EE: **A systematic analysis of Tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development.** *Dev Cell* 2009, **16**(2):280-291.

34.     Salzer CL, Elias Y, Kumar JP: **The retinal determination gene eyes absent is regulated by the EGF receptor pathway throughout development in Drosophila.** *Genetics* 2010, **184**:185-197.

35.     Linsenmayer TF, Gibney E, Igoe F, Gordon MK, Fitch JM, Fessler LI, Birk DE: **Type V collagen: molecular structure and fibrillar organization of the chicken alpha 1(V) NH2-terminal domain, a putative regulator of corneal fibrillogenesis.** *J Cell Biol* 1993, **121**(5):1181-1189.

36.     Ankra-Badu GA, Shriner D, Le Bihan-Duval E, Mignon-Grasteau S, Pitel F, Beaumont C, Duclos MJ, Simon J, Porter TE, Vignal A *et al*: **Mapping main, epistatic and sex-specific QTL for body composition in a chicken population divergently selected for low or high growth rate.** *BMC Genomics* 2010, 11:107.

# 5

# Whole Genome SNP Discovery and Analysis of Genetic Diversity in Turkey (*Meleagris gallopavo*)

Muhammad L Aslam[1*], John WM Bastiaansen[1], Martin G Elferink[1], Hendrik-Jan Megens[1], Richard PMA Crooijmans[1], Le Ann Blomberg[2], Robert C Fleischer[3], Curtis P Van Tassell[4], Tad S Sonstegard[4], Steven G Schroeder[4], Martien AM Groenen[1], Julie A Long[2]

[1] *Animal Breeding and Genomics Centre, Wageningen University, 6709PG, Wageningen, the Netherlands*
[2] *Animal Biosciences and Biotechnology Laboratory, Animal and Natural Resources Institute, Beltsville Agricultural Research Center, United States Department of Agriculture, Beltsville, MD 20705, USA*
[3] *Center for Conservation and Evolutionary Genetics, Smithsonian Conservation Biology Institute, Washington, DC, 20013, USA*
[4] *Bovine Functional Genomics Laboratory, Animal and Natural Resources Institute, Beltsville Agricultural Research Center, United States Department of Agriculture, Beltsville, MD 20705, USA*

## Abstract

### Background

The turkey (*Meleagris gallopavo*) is an important agricultural species and the second largest contributor to the world's poultry meat production. Genetic improvement is attributed largely to selective breeding programs that rely on highly heritable phenotypic traits, such as body size and breast muscle development. Commercial breeding with small effective population sizes and epistasis can result in loss of genetic diversity, which in turn can lead to reduced individual fitness and reduced response to selection. The presence of genomic diversity in domestic livestock species therefore, is of great importance and a prerequisite for rapid and accurate genetic improvement of selected breeds in various environments, as well as to facilitate rapid adaptation to potential changes in breeding goals. Genomic selection requires a large number of genetic markers such as e.g. single nucleotide polymorphisms (SNPs) the most abundant source of genetic variation within the genome.

### Results

Alignment of next generation sequencing data of 32 individual turkeys from different populations was used for the discovery of 5.49 million SNPs, which subsequently were used for the analysis of genetic diversity among the different populations. All of the commercial lines branched from a single node relative to the heritage varieties and the South Mexican turkey population. Heterozygosity of all individuals from the different turkey populations ranged from 0.17-2.73 SNPs/Kb, while heterozygosity of populations ranged from 0.73-1.64 SNPs/Kb. The average frequency of heterozygous SNPs in individual turkeys was 1.07 SNPs/Kb. Five genomic regions with very low nucleotide variation were identified in domestic turkeys that showed state of fixation towards alleles different than wild alleles.

### Conclusion

The turkey genome is much less diverse with a relatively low frequency of heterozygous SNPs as compared to other livestock species like chicken and pig. The whole genome SNP discovery study in turkey resulted in the detection of 5.49 million putative SNPs compared to the reference genome. All commercial lines appear to share a common origin. Presence of different alleles/haplotypes in the SM population highlights that specific haplotypes have been selected in the modern domesticated turkey.

## 5.1 Introduction

All commercial turkey lines descend from the South Mexican turkey (*Meleagris gallopavo gallopavo)* indigenous to Mexico, first domesticated in 800 BC [1]. In the US, the turkey is registered as a single breed with eight different varieties as defined primarily by plumage colour. Five of these eight varieties (Bronze, Narragansett, White Holland, Black and Slate) were registered in 1874, while the remaining three (Beltsville Small White, Bourbon Red, and Royal Palm) were registered in 1951, 1909, and 1971 respectively. There are a total of five wild turkey subspecies in North America that are not genetically related to modern commercial lines [1].

Turkey is the second largest contributor of poultry meat consumed worldwide [2]. The production per bird doubled, largely due to selection pressure by the primary breeders for specific economically important traits, such as body weight, meat quality, and egg production [3-6]. Historically, quantitative genetics-based selection has been the primary strategy of genetic improvement of livestock [7]. This genetic improvement was largely applied to highly heritable traits, such as body size and breast muscle development. Genetic improvement of farm animals through selection may have increased production but has also resulted in a loss of genetic diversity [8]. The efficiency of these classical methods used for genetic improvement decreases when applied to traits that are difficult to measure or have lower heritability [7]. The availability of genome-based selection, based on a large number of SNPs at a density equivalent to the resolution of linkage disequilibrium (LD), has the potential to transform breeding and incorporate previously unavailable genetic information into commercial lines [9] which can be expected to change the impact of commercial breeding on diversity. A tremendous loss of poultry genetic diversity has been observed within research institutions in the United States and Canada over the past 4 decades due to selection in chickens [10]. SNPs are a good marker type to study diversity. SNPs represent the most abundant source of genetic variation within the genome and are linked to heritable differences between individuals [11]. In addition, SNPs have a low mutation rate and are thought to be good genetic markers of potential disease phenotypes as well as for other complex traits[12]. Moreover, SNPs are valuable markers for a variety of genetic and genomic applications such as the construction of genetic and physical maps and the analysis of genetic diversity [13]. Next generation sequencing (NGS) has proven to be very effective for the large scale, genome-wide discovery of this type of genetic variation [14, 15]. When a high quality reference genome sequence is available, genomic sequences of individuals can be aligned more easily to this reference genome to detect nucleotide variation[15, 16]. NGS

97

platforms allow for highly redundant coverage of the genome, a prerequisite for high quality genome-wide SNP discovery in the complex genomes of plants and animals [15, 17, 18].

The genome assembly, containing 39 autosomes and 2 sex chromosomes, of the turkey became available recently[19]. The size of the turkey genome assembly is 1.1 billion bases and, to date, about 600,000 SNPs [15, 19] have been identified within the reference genome assembly. Increasing the number of SNPs identified in the turkey is an essential step for future improvement of economically important traits through genetic association studies [20-22].

Domestication of livestock species and a long history of migrations, selection and adaptation has created an enormous variety in breeds in livestock [8]. Phenotypic selection has created a wide diversity of breeds that are adopted to different climatic conditions and purposes [23]. Phenotypic variation observed between and among breeds of domestic animals is overwhelming compared with that in natural populations [23]. Chicken is considered the most closely related species of the turkey. The observed phenotypic diversity in chicken is much larger than that of turkey, [23, 24] most likely reflecting a much larger effective population size of chicken, before specialized commercial populations were established during the twentieth century. This is consistent with the extensive sequence diversity present in domestic chicken (5 SNPs/Kb) [25, 26].

The presence of genetic diversity in domestic livestock species is of great importance for sustained genetic improvement of selected breeds in various environments, as well as to facilitate rapid adaptation to potential changes in breeding goals [27, 28]. In animal breeding, crosses with non-commercial populations are rarely applied and genetically improved animals are often kept in small, closed populations. Small effective population sizes and epistasis can result in loss of genetic diversity, which can lead to reduced individual fitness and reduced response to selection [29, 30]. Several studies have assessed genetic diversity in different livestock species [29, 31-37] using different types of markers. A number of genetic diversity studies in chicken have reported loss of genetic diversity in commercial chicken populations because of high selection pressure and low effective population size [32, 34, 38]. A few studies have been published that explored genetic diversity in turkey genetic resources. However, these studies used a limited number of molecular markers [39, 40] and only one study has been published that used 9 SNPs along with other molecular markers [41].

The goal of this project was to investigate turkey genome variation and to provide a resource for subsequent genomic work in the turkey and to cover a wide sampling of population for the development of a high-density SNP chip with minimal

ascertainment bias. We have used the identified SNPs to estimate relatedness among the sequenced turkey populations, which will uncover the genetic diversity available to breeders. Information of genetic diversity can be used in the design of breeding programs including making decisions on introgression of novel genes that may affect economically important traits such as growth, meat quality, fitness, and survival traits.

## 5.2 Methods

### Populations

Eleven turkey populations were available for this study. Males from seven commercial lines, three heritage varieties and 113 years old samples of wild turkeys from South Mexico (SM turkeys) were used for whole genome sequencing. The seven commercial lines, L1 through L7, were obtained from two different primary breeding companies. The three heritage varieties were the Beltsville Small White (BvSW), the Royal Palm (RP) and the Narragansett (Nset) [42-44]. Tissue samples representing the wild population were obtained from the Bird Collection of the Smithsonian Institution's National Museum of Natural History (USNM 165490, USNM 166330, and USNM 166329), and were originally collected in 1899 from Chihuahua, Mexico. These samples represent the progenitor subspecies, the South Mexican (SM) turkey. In total 32 individuals were selected for whole genome re-sequencing, with three males per population except for RP, which was represented by 2 males.

### Genomic DNA Extraction, Library Preparation and Sequencing

Considering mature erythrocytes in poultry are nucleated, genomic DNA was extracted from whole blood of the commercial and heritage lines with the QIAamp DNA blood Midi Kit (Qiagen, Valencia, CA); the procedure included a proteinase K digestion followed by column purification. Integrity of high molecular weight DNA following the extraction was confirmed by agarose gel analysis. Genomic DNA was sheared using the Covaris S2 to yield an average fragment size of 450 bp, as determined with the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA). The DNA from the three historic SM samples was extracted from the toe-pads in the ancient DNA laboratory at the Smithsonian Institution's Center for Conservation and Evolutionary Genetics, that is fully equipped to avoid contamination with modern DNA. DNA extraction followed a standard protocol of proteinase k and DTT digestion followed by phenol-chloroform extraction and centrifugal dialysis with Centricon concentrators (following methods provided in [45]). An extraction blank sample was used as a no-sample control in each round of extraction. Extractions involved alternation of turkey samples with samples from other avian or non-avian

taxa, in order to detect potential cross-contamination among extracts. Extracts of the samples and extract controls were subjected to PCR with standard avian mtDNA primer sets (Cytochrome *b*, ND2; [46]) followed by sequencing of positive products to confirm the isolation of turkey DNA from the toe pads. The genomic DNA of the SM samples ranged from 40-43bp (Agilent Bioanalyzer).

Genomic libraries were prepared with the Paired-end Sequencing Sample Preparation Kit (Illumina, San Diego, CA) with 5 μg of genomic DNA for commercial and heritage lines according to the manufacturer's instructions; for the SM samples the molar equivalent of 5 μg was used to construct the libraries. All genomic DNA libraries were validated with the Agilent Bioanalyzer (model 2100). The automated cBot Cluster Generation System (Illumina) was used to generate clusters on the flow cell. Each individual was sequenced (paired-end; read length 120 bp) in a single lane of a flow cell using the Illumina GAIIx. The DNA extracted from museum samples for the SM turkeys was highly degraded, and thus single-end reads of 40 bp were generated from these samples.

### *Sequence mapping and SNP identification*

Sequence reads of each individual from the domesticated populations (heritage varieties and commercial lines) were filtered on base quality; reads were trimmed if three consecutive bases had an average Phred-like quality score of less than 13. Both sequences in a pair needed to exceed 40 bp in length after trimming to be retained for analyses. Sequence reads from the individuals of the SM population were not quality-trimmed before further analyses since they were sequenced to a length of 40bp only. Sequence reads were aligned against the turkey reference genome (UMD 2.01) using the MOSAIK aligner [47]. Mapping of reads from each individual to the reference genome sequence was performed with hash size 15 (hs), 100 maximum hash positions (mhp), an alignment candidate threshold (act) of 20, and a maximum mismatch percentage (mmp) of 5. Banded Smith-Waterman algorithm (bw = 41) was used to increase the speed of alignments. The algorithm implemented in MOSAIK calculates a mapping quality for each sequence and measures the probability that a sequence belongs to a specific target. The alignments were sorted using MosaikSort. Finally, the file was converted to BAM format [48] using MosaikText. All BAM files have been uploaded to NCBI's Sequence Read Archive (SRA) database under the study accession number "SRP012021.2".

The mpileup function of SamTools version 0.1.12a [49] was used to call variants, separately for each turkey population. The view option of bcftools [49] was used to call the genotype at each variant for each animal. Genotypes were called for each animal with a minimum genotype quality of 20, and a read depth between 1 and

25. At least one individual in a population needed to have a genotype call that met these criteria at a particular position. A SNP that passed the above mentioned criteria were considered as a putative SNP. Putative SNPs were categorized into fixed differences compared to the reference genome and segregating SNPs. Homozygous non-reference genotypes that were the same in all individuals of a population were considered fixed SNPs, while the SNPs that had variable/heterozygous genotypes in a population were considered segregating SNPs.

To estimate heterozygosity (heterozygous SNPs/kb), mpileup genotyping analysis (described above) was used and the number of heterozygous SNPs was calculated at the reference bases covered from 5 to 10 fold. For each individual in a population, heterozygosity was estimated by dividing the total number of discovered heterozygous SNPs by the total genome sequence covered from 5 to 10 fold. Population heterozygosity was estimated by averaging the heterozygosity of all individuals within a population.

### *Functional annotation of SNPs*

The gene-based analysis of ANNOVAR software [50] was used to functionally annotate the putative SNPs. For each putative SNP, the location (exonic, intronic, intergenic, 5'UTR, 3'UTR, splice acceptor or donor site, downstream or upstream) and the functional annotation (nonsynonymous, synonymous, stop codon gain or loss, and amino acid changes) were determined based on the turkey reference genome (UMD 2.01). Gene annotations used in this analysis were taken from Ensembl [51]. Standard settings for gene based analysis of ANNOVAR were used.

### *Nucleotide diversity and false discovery rate*

Genome wide mapping density, or read depth distribution, and the nucleotide diversity across the whole genome were assessed for each individual of the 11 turkey populations. Read depth distribution was used to calculate average sequence coverage across the whole genome. To get genotypes of each individual without imputation, pileup function of SamTools version 0.1.12a [49] was used for the estimation of nucleotide diversity across the whole genome. Genotypes were called for each individual using minimum genotype quality of 20, and a read depth between 3 and 15. The number of heterozygous and homozygous non-reference SNP calls was estimated compared to the reference genome within a 300 Kb window. In order to estimate SNP false discovery rate (FDR), 30 large genomic regions of variable sizes (ranging from 2.7-10.5 Mb on variable positions at chromosomes 1, 3 and 10) were investigated where one individual from each of the 10 domesticated populations was clearly homozygous for a single haplotype. Homozygous regions were identified by visual inspection of the nucleotide diversity

plots for turkey chromosome 1, 3 and 10. Any SNP within these regions were considered to be false positives. The false discovery rate was calculated as the total number of heterozygous SNP positions divided by the total number of bases covered (1-25 fold coverage) in these 30 regions.

### Genetic diversity analysis

PHYLIP software, version 3.69 [52] was used to calculate pairwise Nei's genetic distance [53] among all the individuals from the 11 turkey populations. SNPs for which genotypes were called in at least 9 turkey populations (irrespective of whether SNPs were segregating in all these populations) were selected and utilized for the genetic diversity analysis. Threshold of at least 9 turkey population was selected to increase number of selected SNPs for analysis and to make sure presence of selected SNPs in maximum populations to have a reliable genetic comparison. Pairwise genetic distance analyses were based on marker data that the individuals had in common, because PHYLIP is unable to deal with missing data [33]. Mega 5.0 [54] was used for hierarchical clustering using a Neighbour-joining procedure on the genetic distance matrix for all the individuals. The wild population was used to root the phylogenetic tree.

### Non-reference allelic state

The genome of each individual was screened, using the nucleotide diversity analysis described above, for the occurrence of non-reference allelic states. Determining the ancestral allelic state of SNPs was not possible because species with appropriate evolutionary distance are not available. Chicken is considered a closely related species to turkey but the evolutionary distance to the last common ancestor of these two species is around 30 million years [55]. To quantify regional changes in genomic diversity between SM and the domesticated populations, we used heterozygosity as well as the presence of non-reference allelic homozygosity of the positions sufficiently covered by sequencing.

The difference in non-reference allele homozygosity between domesticated and the SM turkey populations was calculated for each bin. This difference was then divided by the average homozygous non-reference allele SNP density for the bin to yield a relative measure that can be compared between bins with different levels of variation.

The ratio of non-reference homozygosity in wild SM vs. domesticated populations was calculated within bin sizes of 300 Kb. A high ratio points to non-reference alleles being lost, or decreased in frequency during domestication and selection. A high ratio of non-reference homozygosity, in combination with low heterozygosity in the domesticated populations, is interpreted as a reduction of allelic variation from wild to domesticated populations, or "fixation of the reference alleles". A bin

was considered "fixed for the reference allelic state" in domesticated populations when two conditions were met. First, bins were considered "fixed" when heterozygosity was equal or lower than 0.0002 on average across all domesticated populations. This threshold was chosen because only 5% of the bins had a heterozysity equal or lower than 0.0002 (1 heterozygous position/5000 bp). Second, bins that were considered "fixed" had to have a ratio of non-reference allele homozygosity above or equal to 1.73, which means that the non-reference allele homozygosity of the wild population must be at least 73% higher than the domesticated populations. This threshold was chosen because only 5% of all the bins in the genome had a ratio equal or higher than 1.73.

### *Ethical approval for the use of animals in this study*

Although animals were used in this study, no direct experiments were performed on them. Blood sample collection was carried out by highly skilled and experienced personnel from the breeding companies. No approval from the ethics committee was necessary according to local legislation.

## 5.3 Results

### *Whole-genome resequencing and SNPs discovery*

The obtained sequence from the DNA samples of the domestic populations (heritage varieties and the commercial lines) varied from 2.30-13.21 Gbp (Giga basepairs) per individual. After quality trimming and alignment of the short reads, the percentage of bases in the reference genome covered by at least 1 and a maximum of 25 reads varied from 47.48 % to 86.13 % for the animals analyzed (Table 5.1). The sequences generated from SM turkeys varied from 0.41-0.82 Gb of sequence per individual. The sequence depth at bases covered by at least one read ranged from 1.38 to 1.81 for the SM samples and 2.07 to 6.72 for the domesticated turkey lines (Table 5.1).

In total, 5.49 million putative SNPs were identified compared to the reference genome (Table 5.2). Of these 5.49 million SNPs, 4.76 million SNPs were segregating in at least one population (Table 5.2). The number of segregating SNPs for the different turkey populations varied from 0.12 to 1.58 million, with the highest number of segregating SNPs observed in L3 and the lowest number observed in SM (Table 5.3). The lowest number of fixed SNPs was observed in L3 and the highest number of fixed SNPs was observed in BvSW (Table 5.3). The transition to transversion (Ti/Tv) ratio of the SNPs discovered is 2.45. Of the total 5.49 million SNPs discovered, 75,254 were located in exonic regions, including 23,795 nonsynonymous , 52,506 synonymous, 377 stop gain and 8 stop loss variants. The

majority of these exonic SNPs, 66,795 or 89% were segregating within the populations analyzed (Table 5.4).

**Table 5.1.** Alignment statistics for the individuals from different turkey populations.

| IDs | Sequence coverage (fold)[1] | Assembly coverage (%)[2] | Assembly coverage 1-25X (%)[3] |
|---|---|---|---|
| **L1a** | 5.12 | 79.04 | 78.93 |
| **L1b** | 4.72 | 83.88 | 84.04 |
| **L1c** | 5.61 | 84.10 | 83.85 |
| **L2a** | 6.54 | 85.91 | 85.85 |
| **L2b** | 6.72 | 86.19 | 86.13 |
| **L2c** | 5.18 | 80.16 | 80.05 |
| **L3a** | 6.32 | 85.98 | 85.68 |
| **L3b** | 5.75 | 85.26 | 85.21 |
| **L3c** | 6.24 | 85.72 | 85.91 |
| **L4a** | 6.19 | 85.58 | 85.51 |
| **L4b** | 5.75 | 84.65 | 84.58 |
| **L4c** | 5.13 | 84.14 | 84.12 |
| **L5a** | 3.52 | 71.18 | 71.14 |
| **L5b** | 5.18 | 71.35 | 71.27 |
| **L5c** | 5.73 | 68.35 | 68.08 |
| **L6a** | 2.88 | 65.14 | 65.13 |
| **L6b** | 4.50 | 77.53 | 77.49 |
| **L6c** | 4.52 | 81.45 | 81.43 |
| **L7a** | 5.46 | 78.59 | 78.39 |
| **L7b** | 4.61 | 57.86 | 57.70 |
| **L7c** | 4.99 | 70.88 | 70.78 |
| **BvSW1** | 4.55 | 83.21 | 83.19 |
| **BvSW2** | 5.72 | 48.33 | 47.48 |
| **BvSW3** | 5.59 | 82.24 | 82.13 |
| **Nset1** | 2.07 | 53.84 | 53.82 |
| **Nset2** | 5.39 | 83.94 | 83.86 |
| **Nset3** | 5.17 | 79.42 | 79.29 |
| **RP1** | 5.31 | 60.31 | 60.05 |
| **RP2** | 5.00 | 63.54 | 63.43 |
| **SMW1** | 1.81 | 47.10 | 47.06 |
| **SMW2** | 1.38 | 29.32 | 29.30 |
| **SMW3** | 1.73 | 45.41 | 45.40 |

[1] Average sequence depth of each base in the reference genome that is covered by at least 1 read. The used turkey reference genome (UMD 2.01) has genome size of 1061982190 bp. [2] Percentage of reference genome that is covered by at least one read. [3] Percentage of reference genome that is covered by 1- 25 reads.

*Heterozygosity*

The number of heterozygous genotypes detected within the individuals from the ten domesticated populations (heritage varieties and the commercial lines) varied from 0.08 to 0.80 million with an average of 0.55 million heterozygous genotypes per individual. Individuals from the SM population showed relatively low numbers of heterozygous SNPs; between 0.01 and 0.07 million.

**Table 5.2.** Heterozygosity and the number of SNP observed in each individual of different turkey populations.

| IDs | Homozygous NR SNP [1] | Heterozygous SNP | Heterozygous SNP 5-10X | Genome covered 5-10X (bp) | Heterozygosity Kb$^{-1}$ |
|---|---|---|---|---|---|
| L1a | 663,406 | 659,351 | 369,849 | 320,663,179 | 1.15 |
| L1b | 686,583 | 648,928 | 385,673 | 396,624,720 | 0.97 |
| L1c | 626,434 | 737,472 | 403,423 | 375,734,398 | 1.07 |
| L2a | 827,249 | 755,318 | 504,787 | 532,961,711 | 0.95 |
| L2b | 896,728 | 757,226 | 514,059 | 554,379,839 | 0.93 |
| L2c | 869,872 | 562,653 | 311,525 | 329,283,144 | 0.95 |
| L3a | 568,439 | 762,252 | 519,228 | 532,049,588 | 0.98 |
| L3b | 434,157 | 427,393 | 567,558 | 527,841,728 | 0.99 |
| L3c | 608,276 | 834,241 | 164,167 | 166,315,925 | 1.08 |
| L4a | 720,530 | 616,567 | 440,086 | 454,905,713 | 0.80 |
| L4b | 760,762 | 692,079 | 385,458 | 439,002,235 | 0.97 |
| L4c | 807,407 | 618,335 | 403,201 | 503,650,627 | 0.88 |
| L5a | 666,287 | 340,436 | 160,698 | 180,577,454 | 0.89 |
| L5b | 652,149 | 352,682 | 165,723 | 144,150,087 | 1.15 |
| L5c | 736,951 | 520,850 | 251,977 | 223,238,275 | 1.13 |
| L6a | 581,773 | 294,736 | 109,405 | 115,435,304 | 0.95 |
| L6b | 644,421 | 567,275 | 330,736 | 306,448,666 | 1.08 |
| L6c | 638,770 | 579,232 | 341,869 | 348,094,277 | 0.98 |
| L7a | 736,881 | 550,299 | 300,174 | 305,785,110 | 0.98 |
| L7b | 698,647 | 379,941 | 185,444 | 161,035,610 | 1.15 |
| L7c | 730,143 | 504,513 | 275,118 | 252,564,184 | 1.09 |
| BvSW1 | 1,053,237 | 417,544 | 241,641 | 372,524,318 | 0.65 |
| BvSW2 | 1,071,513 | 269,338 | 103,333 | 144,219,590 | 0.72 |
| BvSW3 | 1086121 | 525262 | 299,713 | 369,633,525 | 0.81 |
| Nset1 | 643,308 | 79,232 | 25,217 | 144,546,998 | 0.17 |
| Nset2 | 667,797 | 519,815 | 9,929 | 4,717,330 | 2.10 |
| Nset3 | 773,183 | 804,627 | 454,052 | 320,395,210 | 1.42 |
| RP1 | 885,734 | 510,427 | 154,899 | 167,716,001 | 0.92 |
| RP2 | 842,442 | 522,599 | 276,752 | 208,702,070 | 1.33 |
| SMW1 | 551,149 | 69,199 | 11,106 | 9,379,558 | 1.18 |
| SMW2 | 551,380 | 17,275 | 2,030 | 744,899 | 2.73 |
| SMW3 | 551,543 | 44,784 | 6,921 | 6,868,381 | 1.01 |

[1] Homozygous non reference SNPs observed in each individual.

Heterozygosity (heterozygous SNPs/kb) of all individuals from the different turkey populations ranged from 0.17-2.73 while heterozygosity of populations ranged from 0.73-1.64 (Table 5.2 & 5.3). The BvSW population had the lowest heterozygosity, while SM showed the highest heterozygosity within the analyzed populations (Table 5.3). Observed average nucleotide diversity in the 10 largest chromosomes was 0.0005 segregating SNPs per nucleotide position while average nucleotide diversity in the smaller chromosomes (20-30) was 0.0007. Chromosome Z showed the lowest nucleotide diversity with 0.0002 segregating SNPs per nucleotide position. Based on observed homozygous regions (Figure 5.1), interpreted to represent two copies of the same Identical By Descent (IBD) haplotype, the estimated average heterozygous genotype FDR was 0.00002 per nucleotide position in the reference genome (ranging from 0.000012-0.000023 in the different individuals).

**Table 5.3.** Discovered segregating, and the fixed number of SNPs along with the observed heterozygosity Kb$^{-1}$ in each turkey population.

| Population ID | Segregating SNPs[1] | Fixed SNPs[2] | Heterozygosity Kb$^{-1}$ |
|---|---|---|---|
| L1 | 1,563,553 | 617,893 | 1.07 |
| L2 | 1,504,682 | 781,352 | 0.94 |
| L3 | 1,589,525 | 502,807 | 1.01 |
| L4 | 1,441,173 | 709,507 | 0.88 |
| L5 | 950,425 | 674,038 | 1.06 |
| L6 | 1,139,459 | 613,069 | 1.00 |
| L7 | 1,097,788 | 673,807 | 1.07 |
| BvSw | 926,733 | 1,047,010 | 0.73 |
| Nset | 1,194,570 | 708,773 | 1.23 |
| RP | 883,602 | 813,164 | 1.12 |
| SMW | 120,305 | 552,032 | 1.64 |

[1] The total number of SNPs detected compared to the reference genome in which the non-reference allele is segregating in a population. [2] The total number of SNPs detected compared to the reference genome in which only the non-reference allele is found in a population.

**Table 5.4.** Number of SNPs detected.

| Variants | Reference total[1] | Segregating Total[2] |
|---|---|---|
| Nonsynonymous | 23,795 | 20,463 |
| Synonymous | 52,506 | 47,281 |
| Stopgain | 377 | 295 |
| Stoplost | 8 | 7 |
| | | |
| Exonic splice site | 1,437 | 1,256 |
| Exonic | 75,254 | 66,795 |
| Splice acceptor or donor site (interonic) | 734 | 607 |
| 5'UTR/3'UTR | 8,933 | 7,661 |
| Upstream/downstream | 142,829 | 124,005 |
| Intronic | 1,749,427 | 1,518,783 |
| Intergenic | 3,514,102 | 3,044,243 |
| ncRNA | 1,044 | 916 |
| **Total** | **5,493,760** | **4,764,266** |

[1] SNPs detected compared to the reference genome in which the non-reference allele is detected in at least one of the 29 individuals. [2] Detected segregating SNPs within all turkey individuals.

### *Genetic diversity*

There were 223,264 SNPs segregating in at least 9 turkey populations, and these were used to calculate Nei's pair wise genetic distances. The tree based on Nei's genetic distance for the 32 turkey individuals from the 11 different turkey populations presents their genetic relationships (Figure 5.2). Individuals from a specific turkey population clustered closely together. Inter-population comparisons demonstrated that commercial lines formed a cluster that was distinct from heritage lines with the exception of the L5 line, which exhibited a closer genetic relation to the heritage varieties. Among the heritage varieties, RP and Nset were more genetically related than either to BvSW. Individuals from the SM population also clustered together and showed relatively closer genetic relation with BvSW population.
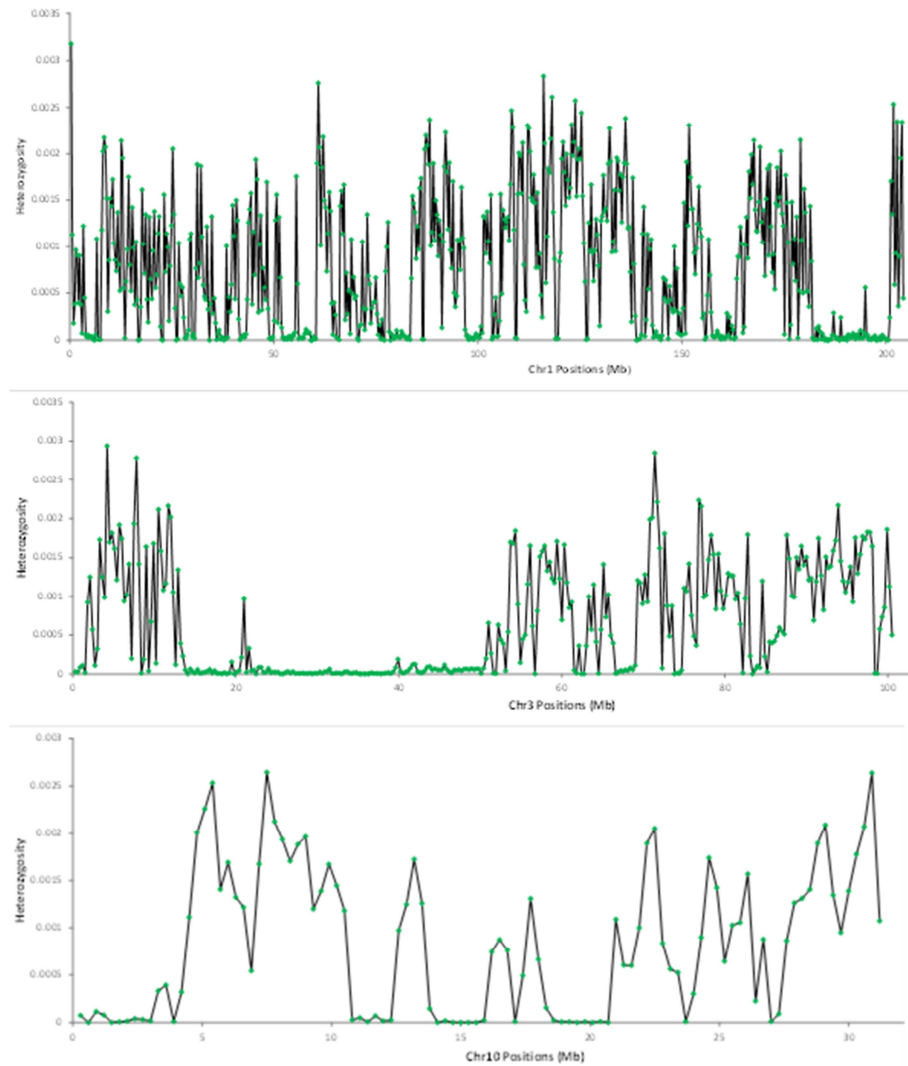
**Figure 5.1:** Nucleotide diversity showing heterozygous and homozygous regions at chromosome 1, 3 and 10. Heterozygosity across chromosome 1, 3 & 10 for L1c individual. On the x-axis chromosome positions (Mb) are presented. On the y-axis heterozygosity is given as density of heterozygous SNPs corrected for the number of bases covered within a window size of 300 Kb. Note the clear homozygous regions at 188-198 Mb for chromosome 1, 24-38 Mb for chromosome 3 and 18-21 Mb for chromosome 3.
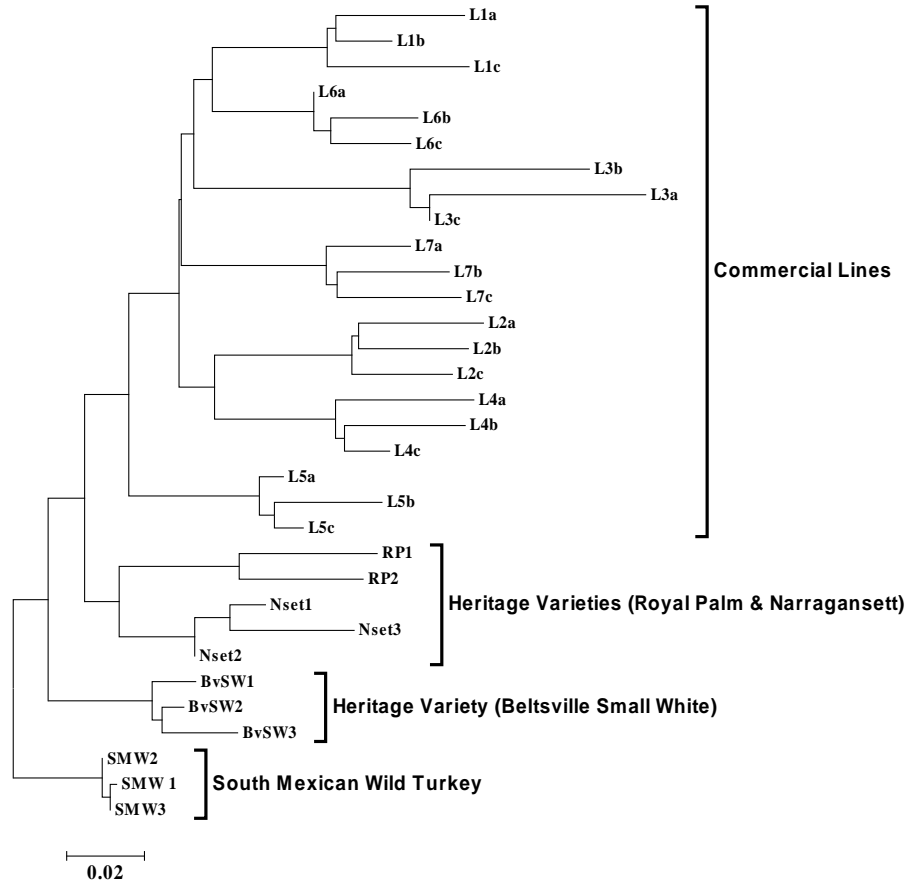
**Figure 5.2.** Dendrogram for 32 individuals from 11 different turkey populations based on Nei's genetic distance.

### *Non-reference allelic state*

Six regions on five different turkey chromosomes (3, 4, 9, 14, and 22) showed differences between the SM and the domesticated populations with respect to the occurrence of no-reference wild type and the reference allelic states (Figure 5.3). Domesticated populations predominantly showed the reference allelic state, while the SM populations predominantly showed the no-reference wild type allelic state within these regions. These six regions were then examined with respect to the heterozygous SNP density per nucleotide positions within the same bin size. Within these six regions, nucleotide diversity for all the domesticated populations was found to be close to zero, except for one region on chromosome 4 that showed high segregation of non-reference alleles within the domesticated populations (Figure 5.4). The other five genomic regions, two regions in chromosome 22 and one region in each of the remaining three chromosomes, (3, 9 and 14), met the criteria mentioned in the methodology section (Additional file 5.1). These genomic regions were considered fixed for the reference allelic state in the domesticated populations.
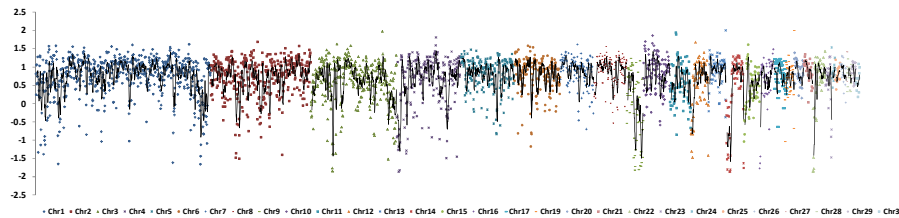


**Figure 5.3:** Difference in non-reference allele density per nucleotide position between domesticated and the wild SM populations. Y-axis denotes difference in non-reference allele density per nucleotide position relative to the mean level of variation discovered between domesticated and the wild SM turkey populations with in a bin size of 300 Kb. Five turkey chromosomes 3, 4, 9, 14 and 22 shows visible difference.
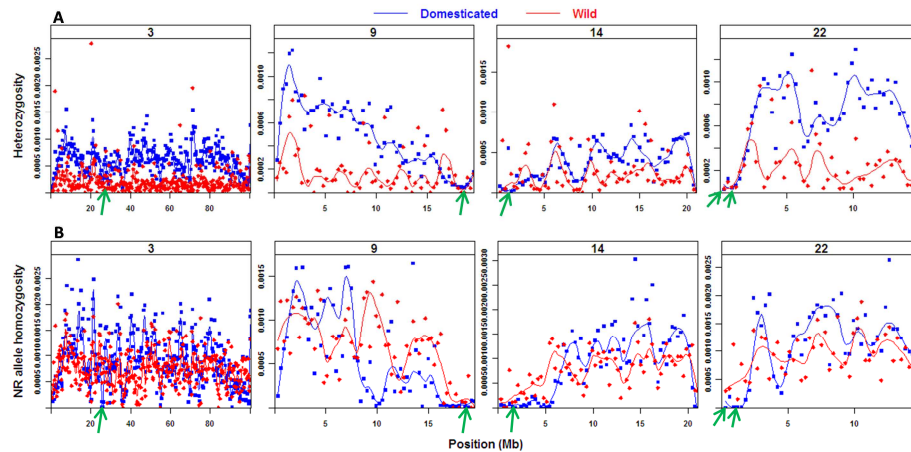
**Figure 5.4:** Comparison of heterozygosity and the non-reference allele homozygosity between wild and domesticated turkeys. **A)** Heterozygous SNP density per nucleotide position (y-axis) within a bin size of 300Kb, x-axis shows positions in million basepairs (Mb) for turkey chromosomes 3, 9, 14 and 22. **B)** Non-reference allele homozygosity per nucleotide position (y-axis) within a bin size of 300Kb, x-axis shows positions in million basepairs (Mb) for turkey chromosomes 3, 9, 14 and 22. Green arrows identify regions fixed for reference haplotype in domesticated populations.

## 5.4 Discussion

In this study, we performed whole genome sequencing for SNPs discovery and used the identified SNPs to characterize genetic diversity in the turkey genome. To avoid imputation of genotype calls across the different populations, mpileup was applied within each population separately because the applied method (mpileup) relies in part on Hardy-Weinberg Equilibrium (HWE) for imputation of genotypes [49].

By using a NGS (Illumina GAIIx) approach, we discovered millions of high quality SNPs in the turkey. Next generation sequencing approaches are considered highly reliable for genome-wide discovery of sequence variation [15], when used to compare different lines/strains to a reference genome [17]. The adoption of NGS platforms for the discovery of genomic variation has now become mainstream [15, 17, 18, 56].

The high quality of the SNPs discovery reported here is reflected by the low FDR of 0.00002 per nucleotide in the genome. This FDR suggests around $2.1 \times 10^4$ false discovered heterozygous positions per turkey genome (size of $1.1 \times 10^9$ base pairs). The SNPs FDR rate for the same 10 animals from distinct turkey populations was estimated after correcting for the coverage and using estimates of FDR per

nucleotide position. The SNPs FDR was found to be 2.6 %, a number that is similar in magnitude as found previously in the human 1000 Genome Project. In addition to the low FDR, we found a transition/transversion (Ti/Tv) ratio within the expected range. The expected Ti/Tv ratio of true novel variants can vary with the targeted region (whole genome, exome, specific genes), species and also can vary greatly by the CpG and GC content of the region [18, 56, 57]. In the case of exomes, an increased presence of methylated cytosine in CpG dinucleotides in exonic regions leads to an increased Ti/Tv ratio [57] due to an easy deamination and transition of a methylated cytosine to a thymine [57]. It is also observed that GC content is higher in birds and mammals than in invertebrates [58]. Observed Ti/Tv ratio in our study of turkey is in concordance with the findings from Dalloul et al. [19], but slightly higher (2.45) than that of human. This higher ratio is most likely explained by the smaller genome size and a higher GC percentage in bird genomes.

We report the number of segregating as well as total number of SNPs with their functional annotation. The 23,795 nonsynonymous variants that were observed can potentially change the structure of proteins, possibly resulting in altered phenotypes [59]. We observed 5,417,069 SNPs that were present in non-protein coding DNA. Furthermore, we discovered 1,749,427 intronic variants, some of which may alter gene expression or result in alternative splicing [60, 61]. Variants located in intergenic regions, such as promoter, enhancer and silencer regions can result in altered gene expression. The human genome comprises over 98% non-protein coding DNA [62]. Estimates suggest that at least 5.5% of the human genome, including 3.5% of its noncoding fraction, consists of regions under purifying natural selection against deleterious alleles [63-65]. In addition, most of the variants involved in complex genetic diseases in humans are not located in coding regions [56]. Likewise, variation outside of coding regions may be responsible for economically important traits in domesticated species, e.g. disease resistance, meat quality, efficient growth, or high egg production. The functional information of these variants can help in prediction of phenotypes or genetic merit with higher accuracy and selection of individuals can be done accordingly.

The estimated average frequency of 1.07 heterozygous SNPs Kb$^{-1}$ in the turkey is substantially lower than in chicken, which was previously reported as 4.28 and 2.24 heterozygous SNPs Kb$^{-1}$ in two different studies [25, 26]. In our study, heterozygous SNP discovery was found to be affected by the sequence coverage (e.g. sequence coverage in L6a, Nset1 and the SM animals was low and as a result the number of observed heterozygous SNPs was also low). Estimates of heterozygosity were therefore obtained only from genomic regions that were covered 5 to 10X to adjust for the effect of low sequence coverage.

Modern commercial turkey lines are derived from historic turkey populations that displayed low variation as a result of small effective population size [66, 67]. Heritage (Nset and RP) and the wild SM turkey populations showed higher heterozygosity compared to the commercial populations, which is concordant with the findings of previous studies on ancient and overexploited species [68-70]. The heritage variety BvSW showed the lowest heterozygosity of all turkey populations, which is consistent with the severe bottleneck that this population went through in 2000 (Alexandra Scupham, Personal communications).

Most birds have a characteristic division in chromosome size, with 5 or 6 large chromosomes, around 5 intermediate size chromosomes, and 25 to 30 very small chromosome pairs. In our study, we observed higher nucleotide diversity on smaller chromosomes compared to the larger turkey chromosomes which is in agreement with the previous study [71]. Since the recombination rate is far higher at the smaller sized turkey chromosomes as compared to large chromosomes [72], which leads to lower linkage disequilibrium and higher haplotype diversity on the smaller chromosomes [73]. Although the high gene-density of the smaller chromosomes would make them susceptible to hitchhiking effects that could erode genetic variation, hitchhiking effects appear to be offset by the far higher recombination rate of the micro-chromosomes. Chromosome Z showed the lowest nucleotide diversity, which is concordant with the findings of Dalloul et al. [19]. This low nucleotide diversity of chromosome Z is likely the result of a lower effective population size of this chromosome and lower recombination rate [74].

The presence of different allelic states in the wild SM and the domesticated populations is a demonstration of their divergence during the course of domestication event. Domesticated turkey lines were selected (artificially or naturally) for non-wild type alleles. Domestication has involved the selection on a desired trait(s) [75], and previous studies on domesticated animals have demonstrated selective pressures on genes related to growth [60] and coat colour [76, 77]. Such studies have also demonstrated that artificial selection might have contributed to reduced polymorphism levels and increased LD in domesticated species [78-81]. On-going directional selection causes footprints of selection identifiable as regions where the derived allele frequency is higher than non-selected regions [26, 82, 83]. Most of the turkey chromosomes are acrocentric and the five genomic regions that were found to be fixed for the reference alleles within the domesticated populations seem to be located close to the centromere [84]. This may explain the presence of a strong hitchhiking effect due to the low recombination rate close to the centromeres. These fixed turkey genomic regions were then investigated for the presence of report QTLs corresponding to these

regions. While QTLs were not found within the fixed regions [85], there were QTLs for growth and meat quality on chromosome 3, a QTL for percentage drip loss on chromosome 14 and a growth related QTL on the chromosome 22 [85]. These QTLs for different traits on chromosomes 3, 14 and 22 were located at distinct positions that did not coincide with the observed regions with high reference allele frequency. Due to the evidence of the presence of structural and functional conservation in the turkey and the chicken genomes [72, 85] and also the limited availability of information on turkey QTLs, these 5 turkey genomic regions that were found to be fixed for reference alleles within domesticated populations, were aligned with the chicken genome sequence (WASHUC2) to determine the position of these turkey genomic regions within the chicken genome (Additional file 5.1). Regions of the chicken genome exhibiting synteny with turkey were then examined for the presence of known chicken QTLs [86]. Several QTL were identified within these 5 genomic regions (Additional File 5.1) and most were related to growth traits (Additional File 5.1). Production census of turkeys from the last few decades [2] show that turkeys are highly selected for growth and this high selection pressure might have favoured reference alleles in domesticated populations. Since several of the regions identified in this study are probably close to a centromere, the effect of selection may have extended over a larger region due to the likely reduced recombination rate in centromeric parts of the genome.

The genetic diversity analysis among the 11 different turkey lines showed that the heritage varieties and the commercial populations are derived from the wild South Mexican population. All of the heritage varieties (BvSW, RP and Nset) are closely related which is in agreement with previously published data [40, 41]. The relatedness of these heritage varieties can probably be explained either by historic nature, a common origin, selection for similar traits/phenotype or a relatively low selection pressure in these varieties. The Nset, RP and BvSW heritage lines were developed in America in 1800, 1920 and 1930, respectively [66, 67]. It is assumed that the colour pattern of RP is derived from crossbreeding with Narragansett and perhaps another variety, as Nset colour mutation is a component of the final RP colour (Smith et al., 2005). The close genetic relatedness observed between RP and Nset in our study is also concordant with that assumption and with previous studies [40, 41]. The close relatedness of the L5 commercial line to the heritage lines is not surprising as it represents a female line selected for medium weight, conformation and egg production; selected traits characteristic of the heritage lines [67]. The other commercial lines that cluster separate from L5 in the dendrogram were selected for different objectives such as higher body weight and rapid growth.

114

## 5.5 Conclusion

The turkey genome is much less diverse with a relatively low frequency of heterozygous SNPs as compared to other livestock species like chicken and pig. The whole genome SNP discovery study in turkey resulted in the detection of 5.49 million putative SNPs compared to the reference genome. All commercial lines appear to share a common origin. Presence of different alleles/haplotypes in the SM population highlights that specific haplotypes have been selected in the modern domesticated turkey.

## Authors' contributions

MLA, JWMB, MGE and HJM analysed the data. LAB assisted and trained laboratory personnel in genomic DNA isolation from blood and preparation of all libraries. RCF isolated the museum specimen DNA. CPVT assisted with project design and coordination. TSS helped develop sequencing strategy and guided library preparation of ancient DNA samples. SGS was responsible for sequencing processing. JAL conceived and developed the project, organized blood sample collection and DNA sequencing, and was the Principal Investigator (Agriculture and Food Research Initiative Competitive Grant no. 2010-65205-20428). MLA wrote the paper and all other authors gave suggestions and comments for the improvement of paper. All authors read and approved the final manuscript. Overall coordination of the project was by JAL, JWMB, RPMAC and MAMG.

## Acknowledgements

## Additional files:

*Additional file 5.1*

**Link:** http://vps6371.xlshosting.net/Aslam/

**Description:**

**Title: Positions of turkey genomic regions with their mapping positions and underling QTL in chicken genome.** This file contains the start and the end positions of turkey genomic regions that showed fixed haplotype for the reference alleles in domesticated populations. This file also contains information about the start and the end positions of these turkey genomic regions in chicken genome and the chicken QTL reported within these regions.

## References

1.    Speller CF, Kemp BM, Wyatt SD, Monroe C, Lipe WD, Arndt UM, Yang DY: **Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication**. *PNAS* 2010, **107**(7):2807-2812

2.    **Food and agriculture organization statistical division (FAOSTAT) of the United Nations** [http://faostat.fao.org/]

3.    Berri C, Wacrenier N, Millet N, Le Bihan-Duval E: **Effect of selection for improved body composition on muscle and meat characteristics of broilers from experimental and commercial lines**. *Poult Sci* 2001, **80**:833-838.

4.    Li H, Deeb N, Zhou H, Mitchell AD, Ashwell CM, Lamont SJ: **Chicken quantitative trait loci for growth and body composition associated with transforming growth factor-β genes1**. *Poult Sci* 2003, **82**:347-356.

5.    Le Bihan-Duval E, Debut M, Berri CM, Sellier N, Santé-Lhoutellier V, Jégo Y, Beaumont C: **Chicken meat quality : genetic variability and relationship with growth and muscle characteristics**. *BMC Genetics* 2008, **9**:53.

6.    Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Ducro BJ, Vereijken A, Megens H-J, Groenen MAM: **Genetic variances, heritabilities and maternal effects on body weight, breast meat yield, meat quality traits and the shape of the growth curve in turkey birds**. *BMC Genetics* 2011, **12**(1):14.

7.    Montaldo HH, Meza-Herrera CA: **Use of molecular markers and major genes in the genetic improvement of livestock**. *EJB* 1998, **1**(2).

8.    Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, Negrini R, Finlay EK, Jianlin H, Groeneveld E *et al*: **Genetic diversity in farm animals - a review**. *Animal Genetics* 2010, **41**:6-31.

9.      Dekkers JCM, Hospital F: **The use of molecular genetics in the improvement of agricultural populations**. *Nature Reviews Genetics* 2002, **3**:22-32.

10.     Fulton JE: **Genomic selection for poultry breeding**. *Animal Frontiers* 2012, **2**(1):30-36.

11.     Suh Y, Vijg J: **SNP discovery in associating genetic variation with human disease phenotypes**. *Mutat Res* 2005, **573**(1-2):41-53.

12.     Gray IC, Campbell DA, Spurr NK: **Single nucleotide polymorphisms as tools in human genetics**. *Human Molecular Genetics* 2000, **9**(16):2403-2408.

13.     Duran C, Appleby N, Edwards D, Batley J: **Molecular genetic markers: Discovery, applications, data storage and visualisation**. *Current Bioinformatics* 2009, **4**:16-27.

14.     Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES: **An SNP map of the human genome generated by reduced representation shotgun sequencing**. *Nature* 2000, **407**(6803):513-516.

15.     Kerstens HHD, Crooijmans RPMA, Veenendaal A, Dibbits BW, Chin-A-Woeng TFC, Dunnen JTd, Groenen MAM: **Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey**. *BMC Genomics* 2009, **10**:479.

16.     Li G, Ma L, Song C, Yang Z, Wang X, Huang H, Li Y, Li R, Zhang X, Yang H *et al*: **The YH database: the first Asian diploid genome database**. *Nucleic Acids Res* 2009, **37**:1025-1028.

17.     Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W *et al*: **Whole-genome sequencing and variant discovery in *C. elegans***. *Nature Methods* 2008, **DOI:10:1038**.

18.     DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al*: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nature Genetics* 2011, **43**(5):491-498.

19.     Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Ann Blomberg L, Bouffard P, Burt DW, Crasta O, Crooijmans RPMA *et al*: **Multi-Platform Next-Generation Sequencing of the Domestic Turkey (*Meleagris gallopavo*): Genome Assembly and Analysis**. *PLoS Biol* 2010, **8**(9):e1000475.

20.     Ma L, Runesha HB, Dvorkin D, Garbe JR, Da Y: **Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies**. *BMC Bioinformatics* 2008, **9**:315.

21.     Boschiero C, Rosario MF, Ledur MC, Campos RLR, Ambo M, Coutinho LL, Moura ASAMT: **Associations between microsatellite markers and traits related to performance, carcass and organs in chickens**. *Int J Poult Sci* 2009, **8**(7):615-620.

22.     Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, Li H, Chen K, Ye S, Ge C *et al*: **Genome-wide association study of body weight in chicken F2 resource population**. *PLoS ONE* 2011, **6**(7):e21872.

23.     Andersson L: **Genetic dissection of phenotypic diversity in farm animals**. *Nature Rev Genet* 2001, **2**:130-138.

24.     American Poultry Association: **APA recognized breeds and varieties**. In*.*; 2010.

25.     International Chicken Polymorphism Map Consortium: **A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms.** *Nature* 2004, **432**:717-722.

26.     Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al*: **Whole-genome resequencing reveals loci under selection during chicken domestication**. 2010, **464**(7288):587-591.

27.     Notter DR: **The importance of genetic diversity in livestock populations of the future**. *J Anim Sci* 1999, **77**:61-69.

28.     Bijma P, Meuwissen THE, Woolliams JA: **Design of sustainable breeding programs in developed countries**. In: *7th World Congress on Genetics Applied to Livestock Production: 2002*; 2002: CDROM paper 24–01.

29.     Ryman N, Utter F, Laikre L: **Protection of intraspecific biodiversity of exploited fishes. Rev. Fish Biol. Fish. 5, 417–446**. *Rev Fish Biol Fish* 1995, **5**:417-446.

30.     Ficetola GF, Garner TW, De Bernardi F: **Genetic diversity, but not hatching success, is jointly affected by postglacial colonization and isolation in the threatened frog, Rana latastei**. *Mol Ecol* 2007, **16**(15):3285.

31.     Kijowski J, Niewiarowicz A: **Emulsifying properties of proteins and meat from broiler breast muscles as affected by their initial pH values**. *J Food Technol* 1978, **13**:451-459.

32.     Hillel J, Granevitze Z, Twito T, Ben-Avraham D, Blum S, Lavi U, David L, Feldman MW, Cheng H, Weigend S: **Molecular markers for the assessment of chicken biodiversity**. *World's Poultry Science Journal* 2007, **63**:33-45.

33.     Megens HJ, Crooijmans RP, San Cristobal M, Hui X, Li N, Groenen MA: **Biodiversity of pig breeds from China and Europe estimated from pooled**

DNA samples: differences in microsatellite variation between two areas of domestication**. *Genet Sel Evol* 2008, **40**(1):103-128.

34.     Muir WM, Wong SG, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, Megens H-J, Zhang H, Okimoto R, Vereijken A *et al*: **Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds**. *PNAS* 2008, **105**(45):17312–17317.

35.     Kijas JM, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S *et al*: **A genome wide survey of SNP variation reveals the genetic structure of sheep breeds**. *PLoS ONE* 2009, **4**(3):e4668.

36.     Gautier M, Laloë D, Moazami-Goudarzi K: **Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds**. *PLoS One* 2010, **5**(9):pii: e13038.

37.     Lin BZ, SasazaKi S, Mannen H: **Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers**. *Ani Sci J* 2010, **81**(3):281–289.

38.     Sawai H, Kim HL, Kuno K, Suzuki S, Gotoh H, Takada M, Takahata N, Satta Y, Akishinonomiya F: **The Origin and Genetic Variation of Domestic Chickens with Special Reference to Junglefowls *Gallus g. gallus* and *G. varius***. *PLoS One* 2010, **5**(5):e10639.

39.     Mock KE, Theimer TC, Rhodes Jr OE, Greenberg DL, Keim P: **Genetic variation across the historical range of the wild turkey (*Meleagris gallopavo*)**. *Molecular Ecology* 2002, **11**:643-657.

40.     Kamara D, Gyenai KB, Geng T, Hammade H, Smith EJ: **Microsatellite marker-based genetic analysis of relatedness between commercial and heritage turkeys (*Meleagris gallopavo*)**. *Poult Sci* 2007, **86**:46-49.

41.     Smith EJ, Geng T, Long E, Pierson FW, Sponenberg DP, Larson C, Gogal R: **Molecular analysis of the relatedness of five domesticated turkey strains**. *Biochemical Genetics* 2005, **43**:35-47.

42.     **The American Livestock Breeds Conservancy; Turkeys: Narragansett**

43.     Marsden SJ: **The Beltsville small white turkey**. In. Beltsville Maryland, USA: Animal husbandry research division: 32-41.

44.     **The American Livestock Breeds Conservancy; Turkeys: Royal Palm**

45.     Slikas B, Jones IB, Derrickson SR, Fleischer RC: **Phylogenetic relationships of insular Micronesian white-eyes (Aves: *Passeriformes: Zosteropidae*), based on mitochondrial sequence data**. *Auk* 2000, **117**:355-365.

46.     Fleischer RC, Kirchman JJ, Dumbacher JP, Bevier L, Dove C, Rotzel NC, Edwards SV, Lammertink M, Miglia K, S. MW: **Mid-Pleistocene divergence of Cuban and North American ivory-billed woodpeckers**. *Biology Letters* 2006, **2**:466-469.

47.     Stromberg M: **Mosaik Assembler**. In: *110014.* Edited by Lee W-P: Boston College; 2010.

48.     The SAM Format Specification Working Group: **The SAM format specification**. In*.*, 1.4-r985 edn; 2011.

49.     Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.

50.     Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data**. *Nucleic Acids Res* 2010, **38**(16):e164.

51.     **Ensembl Genome Browser**. In*.*

52.     Felsenstein J: **PHYLIP (Phylogeny Inference Package)**. In*.*, 3.69 edn. Washington: Department of genome sciences, university of Washington, Seattle; 2005.

53.     Nei M: **Genetic distance between populations**. *American Naturalist* 1972, **106**:283-292.

54.     Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods**. *Molecular Biology and Evolution* 2011, **10.1093**.

55.     Pereira SL, Baker AJ: **A molecular timescale for galliform birds accounting for uncertainty in time estimates and heterogeneity of rates of DNA substitutions across lineages and sites**. *Molecular Phylogenetics and Evolution* 2006, **38**(2):499-509.

56.     Consortium TGP: **A map of human genome variation from population-scale sequencing**. *Nature* 2010, **467**:1061.

57.     Rosenberg MS, Subramanian S, Kumar S: **Patterns of transitional mutation biases within and among mammalian genomes**. *Mol Biol Evol* 2003, **20**(6):988-993.

58.     Bernardi G: **Isochores and the evolutionary genomics of vertebrates**. *Gene* 2000, **241**(1):3-17.

59.     Strachan T, Read AP: **Human Molecular Genetics**. In: *An overview of mutation, polymorphism, and DNA repair.* 2nd edn. New York: Wiley-Liss; 1999.

60. Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M *et al*: **A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig**. *Nature* 2003, **425**:832-836.

61. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes**. *Nature* 2008, **456**(7221):470-476.

62. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al*: **The Sequence of the Human Genome**. *Science* 2001, **291**(5507):1304-1351.

63. Mouse Genome Sequencing Consortium: **Initial sequencing and comparative analysis of the mouse genome**. 2002, **420**(6915):520-562.

64. Miller W, Makova KD, Nekrutenko A, Hardison RC: **Comparative genomics** *Annual Review of Genomics and Human Genetics* 2004, **5**:15-56.

65. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E *et al*: **A high-resolution map of human evolutionary constraint using 29 mammals**. 2011, **478**(7370):476-482.

66. Sponenberg DP, Hawes RO, Johnson P, Christman CJ: **Turkey conservation in the United States**. *Animal Genetic Resources Information Bulletin (AGRI)* 2000, **27**:59-66.

67. Frank R, Reese J, Bender M, Sponenberg DP, Williamson D, Beranger J: **Selecting your best turkeys for breeding**. In*.*: The American Livestock Breeds Conservancy.

68. Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, Carvalho GR: **Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*)**. *Proc Natl Acad Sci* 2002, **99**:11742–11747.

69. Larson S, Jameson R, Bodkin J, Staedler M, Bentzen P: **Microsatellite DNA and mitochondrial DNA variation in remnant and translocated sea otter (*Enhydra lutris*) populations**. *J Mammal* 2002, **83**:893-906.

70. Nabata D, Masuda R, Takahashi O, Nagata J: **Bottleneck effects on the sika deer Cervus nippon population in Hokkaido, revealed by ancient DNA analysis**. *Zoolog Sci* 2004, **21**:473–481.

71. Axelsson E, Webster MT, Smith NGC, Burt DW, Ellegren H: **Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes**. *Genome Res* 2005, **15**(1):120-125.

72. Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Vereijken A, Megens H-J, Groenen MAM: **A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes**. *BMC Genomics* 2010, **11**:647.

73. Megens HJ, Crooijmans RP, Bastiaansen JW, Kerstens HH, Coster A, Jalving R, Vereijken A, Silva P, Muir WM, Cheng HH *et al*: **Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken**. *BMC Genetics* 2009, **10**(86).

74. Ellegren H: **Molecular evolutionary genomics of birds**. *Cytogenetic and Genome Research* 2007, **117**(1-4):120-130.

75. Price EO: **Behavioral development in animals undergoing domestication** *Applied Animal Behaviour Science* 1999, **65**:245-271.

76. Moller M, Chaudhary R, Hellmén E, Höyheim B, Chowdhary B, Andersson L: **Pigs with the dominant white coat color phenotype carry a duplication of the *KIT* gene encoding the mast/stem cell growth factor receptor**. *Mammalian Genome* 1996, **7**(11):822-830.

77. Fang M, Larson G, Soares Ribeiro H, Li N, Andersson L: **Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs**. *PLoS Genet* 2009, **5**(1):e1000341.

78. Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M: **Extensive genome-wide linkage disequilibrium in cattle**. *Genome Res* 2000, **10**:220-227.

79. McRae AF, McEwan JC, Dodds KJ, Wilson T, Crawford AM, Slate J: **Linkage disequilibrium in domestic sheep**. *Genetics* 2002, **160**(3):1113-1122.

80. Amaral AJ, Megens H-J, Crooijmans RPMA, Heuven HCM, Groenen MAM: **Linkage disequilibrium decay and haplotype block structure in the Pig**. *Genetics* 2008, **179**(1):569-579.

81. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, Megens H-J, Zhang H, Okimoto R, Vereijken A *et al*: **Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds**. *PNAS* 2008, **105**(45):17312-17317.

82. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: **A genome-wide scan for signatures of recent selection in Holstein cattle**. *Animal Genetics* 2010, **41**:377-389.

83. Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MAM: **Genome-Wide Footprints**

**of Pig Domestication and Selection Revealed through Massive Parallel Sequencing of Pooled DNA**. *PLoS One* 2011, **6**(4):e14782.

84.    Lundberg M, Kesson SA, Bensch S: **Characterization of a divergent chromosome region in the willow warbler Phylloscopus trochilus using avian genomic resources**. *J Evol Biol* 2011, **24**:1241-1253.

85.    Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Vereijken A, Groenen MAM: **Whole genome QTL mapping for growth, meat quality and breast meat yield traits in turkey**. *BMC Genetics* 2011, **12:61**.

86.    Hu Z-L, Reecy JM: **Animal QTLdb: Beyond a Repository - A Public Platform for QTL Comparisons and Integration with Diverse Types of Structural Genomic Information**. *Mammalian Genome* 2007, **18**:1-4.

# 6

# Genome-wide signatures of selection revealed through massive parallel sequencing of DNA across ten turkey populations

Muhammad L Aslam[1], John WM Bastiaansen[1], Hendrik-Jan Megens[1], Richard PMA Crooijmans[1], Le Ann Blomberg[2], Curtis P. Van Tassell[3], Tad S. Sonstegard[3], Steven G. Schroeder[3], Martien AM Groenen[1], Julie A. Long[2]

[1] Animal Breeding and Genomics Centre, Wageningen University, 6709PG, Wageningen, the Netherlands

[2] Animal Biosciences and Biotechnology Laboratory, Animal and Natural Resources Institute, Beltsville Agricultural Research Center, United States Department of Agriculture, Beltsville, MD 20705, USA

[3] Bovine Functional Genomics Laboratory, Animal and Natural Resources Institute, Beltsville Agricultural Research Center, United States Department of Agriculture, Washington, DC, 20013, USA

## Abstract

### Background

Turkey (*Meleagris gallopavo*, MGA) is an important agricultural domesticated species that is largely used as a meat type bird. Genetic variation in domesticated species and the link of these variation patterns with the events of domestication, selective breeding and the process of evolution is critical for the general understanding of genomic evolution of these species. Selective breeding is expected to leave signatures in the genome of domesticated species, for instance unusually low nucleotide diversity or the presence of exceptionally long homozygous haplotypes. The variation in domesticated species can thus be highly useful not only to gain a better understanding of consequences of selective breeding, but can also aid in elucidating biological and molecular pathways.

### Results

We observed 54 genomic regions that showed significant ($P < 0.05$) signatures of selection on 14 different chromosomes in multiple turkey populations. Areas with evidence of selective sweeps varied from 1.5 Mb to 13.8 Mb in length. Out of these 54 selective sweep regions, 31 were population specific and were observed on 12 different turkey chromosomes while 23 were observed as overlapping regions in multiple populations distributed over 13 different turkey chromosomes. Out of the 31 population specific regions, 26 were present in commercial populations.

### Conclusion

The genome of commercial turkeys showed large selective sweep regions. The relatively high number of sweep regions in commercial populations in comparison to heritage varieties, and the enrichment of turkey sweep regions with genes of importance to growth, indicates that the turkey sweep regions are likely the result of intensive selection for growth moving specific haplotypes towards fixation.

## 6.1 Introduction

Insight in the overall degree of genetic variation in domesticated species and linking these variation patterns to domestication, selective breeding and evolution is critical for the general understanding of genome evolution in these species. Selective breeding is expected to leave changes/signatures in the genome of domesticated species, for instance unusually low nucleotide diversity or the presence of exceptionally long haplotypes [1-3]. Genome-wide characterization of a large number of different breeds and populations for these signatures of selection, along with the functional knowledge of the region can reveal which genes are linked to traits or diseases with a complex genetic basis [4]. The study of variation in domesticated species can thus be highly useful not only to gain a better understanding of the consequences of selective breeding, but also to aid in elucidating biological and molecular pathways [5, 6].

Turkey (*Meleagris gallopavo*, MGA) is an important agricultural domesticated species that is largely used as a meat type bird. All domesticated turkeys, descend from the wild turkeys indigenous to North and South America. There are seven subspecies of the wild form [7] distinguished by geographic range and plumage differences: Mexican (*M. g. gallopavo*), Rio Grande (*M. g. intermedia*), Merriam's (*M. g. merriami*), Gould's (*M. g. mexicana*), Eastern (*M. g. silverstris*), Moore's (*M. g. oneusta*) and Florida (*M. g. osceola*). Three of the seven are purported to play an important role in domestication. It is generally accepted that the first ancestor of the domestic turkey was a Mexican subspecies [8]. The earliest signs of turkey domestication dates to 100 BC-100 AD, at Maya sites such as Cobá [9]. Domestic turkey stocks were established by at least 180 AD within the Tehuacán valley [10], with the South Mexican turkey (*Meleagris gallopavo gallopavo*) as the assumed wild progenitor [8]. Although the wild progenitor of domestic birds of Southwest United States has been long debated, the first strong archaeological evidence for domestic stocks in the Southwest dates to approximately the same time (ca. 200 BC-AD 500 [11]. Domestic turkey has been recognized by the American Standard of Perfection since 1971 [12] and is registered as a single breed with eight different varieties as defined primarily by plumage color. Out of these eight heritage turkey varieties, five (Bronze, Narragansett, White Holland, Black and the Slate) were registered [12] in 1874 while the remaining three (Beltsville small white, Bourbon Red, and the Royal Palm) were registered in 1951, 1909, and 1971 respectively [12]. These domestic turkeys are presumed to be highly inbred [12], and have undergone intensive selection for traits of economic importance such as body weight and meat quality [9, 11].

Recent census shows that turkey is the second largest contributor in worldwide poultry meat production [13]. Global production of turkeys has experienced a massive change and growth over the past 40 years. In 2008, turkey represented 6.65% of the world poultry meat production [4]. Global turkey stocks nearly tripled from 178 million in 1970 to over 482 million in 2008 [4]. Astonishingly, in those four decades, the production volume per bird doubled from 6.74 to 12.66 Kg [4], giving an indication of the a scale of intensive selection in turkeys.

An important genomic indicator of a selective sweep involves local reduction in variation within a selected gene and its adjacent regions[14]. Selection affects the genomic variability which is present in the genome as a diverse array of variants including single nucleotide polymorphisms (SNPs), microsatellites and several types of structural variations (SVs) e.g. large insertions-deletions, inversions, duplications and balanced or unbalanced inter-chromosomal translocations. Next generation sequencing (NGS) is an efficient approach for the large scale, genome-wide SNP discovery and genotyping of individuals [15, 16]. The availability of a high quality reference genome sequence [17] and resequencing of individuals with appropriate genome coverage are key prerequisites for whole genome SNP discovery [15, 16]. Genomic sequences of individuals are aligned to a reference genome to detect nucleotide variations/ differences in genotype of individuals at specific position of genome [18, 19].

Our search was aimed at finding genomic regions where selection or domestication has changed the frequency of favorable alleles towards fixation. The genomic regions that we identified elucidate the effects from the selective pressures or domestication that were applied to turkey.

## 6.2 Methods

### Populations

Ten turkey populations, seven commercial lines and three heritage varieties, were used for whole genome sequencing. The seven commercial lines, L1 through L7, were provided by two different breeding companies. The three heritage varieties were Beltsville Small White (BvSW), Royal Palm (RP) and Narragansett (Nset)[20-22]. In total 29 individuals were selected for whole genome resequencing, with three individuals per population except for RP, which was represented by 2 individuals.

### Genomic DNA Extraction, Library Preparation and Sequencing

Genomic DNA was extracted from whole blood with the QIAamp DNA blood Midi Kit (Qiagen, Valencia, CA); the procedure included a proteinase K digestion followed by column purification. Integrity of high molecular weight DNA following

the extraction was confirmed by agarose gel analysis. Genomic DNA was sheared using the Covaris S2 to yield an average fragment size of 450 bp, as determined with the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA).

Genomic libraries were prepared with the Paired-end Sequencing Sample Preparation Kit (Illumina, San Diego, CA) with 5 µg of genomic DNA according to the manufacturer's instructions. All genomic DNA libraries were validated with the Agilent Bioanalyzer (model 2100). The automated cBot Cluster Generation System (Illumina) was used to generate clusters on the flow cell. Each individual was sequenced (paired-end; read length 120 bp) in a single lane of a flow cell using the Illumina GAIIx.

### Sequence mapping

Sequence reads of each individual turkey were filtered on base quality, where reads were trimmed when three consecutive bases had an average quality score of less than 13. Both sequences in a pair needed to be longer or equal to 40bp after trimming to be retained for analyses. Remaining reads were aligned against the turkey reference genome (UMD 2.01) using MOSAIK aligner [23]. Mapping of reads from each individual to the reference genome sequence was performed with hash size 15 (hs), 100 maximum hash positions (mhp), an alignment candidate threshold (act) of 20 and a maximum mismatch percentage (mmp) of 5. Banded Smith-Waterman algorithm (bw = 41) was used to increase the speed of alignments. The algorithm implemented in MOSAIK calculates a mapping quality for each sequence that measures the probability that a sequence belongs to a specific target. The alignments were filtered for ambiguously mapped reads, and sorted, using MosaikSort. Finally, the file was converted to BAM format [16] using MosaikText. All BAM files have been uploaded to NCBI's Sequence Read Archive (SRA) database under the study accession number SRP012021.2.

### Heterozygosity

Genome wide nucleotide diversity across the whole genome was assessed for each individual of the different turkey populations. The pileup function of SamTools version 0.1.12a [15] was used to estimate nucleotide diversity of each individual across the whole genome. Nucleotide diversity was estimated by calculating the number of heterozygous SNP as well as the number of homozygous non-reference genotypes within each 300Kb window. For calling SNPs, coverage per base was limited to 5-10 fold. Observed number of heterozygous SNPs per nucleotide position were then averaged for each population within the window size of 300Kb.

### Estimation of threshold values for calling selective sweep

Turkey chromosomes were divided into bin sizes of 300 Kb, and these bins were used to estimate threshold values to call selective sweep regions in the genome.

129

Patterns of heterozygosity distribution among these bins were investigated for each turkey population separately. A sweep region was defined when heterozygosity was below the threshold for at least 5 consecutive bins. To obtain genome wide significance thresholds (P < 0.05), heterozygosity values of the bins were randomly permuted across the genome. Subsequently the maximum heterozygosity values from each set of five consecutive bins (each with 300Kb size) was recorded by a sliding window with steps of one bin for the whole genome. In this way, we calculated a threshold for each set of five bins in each replicate of the permutation. Subsequently the lowest threshold of each of 7000 replicates was retained and a 5% threshold was obtained. The 5% threshold heterozygosity value was determined such that for each population we would have a 5% chance of finding 1 sweep region by chance. A threshold of five consecutive bins was used because preliminary results had shown large regions of homozygosity in the turkey genome, and also to obtain stable statistics for heterozygosity. Using these threshold values, each turkey population was investigated for regions of low heterozygosity indicative of the presence of a selective sweep. Subsequently, turkey populations were compared with each other for the overlap in putative sweep regions. Overlapping selective sweep regions were identified when a sweep was replicated in more than one population. The overlapping selective sweep regions were defined as the genomic region covered by the sweeps from at-least two populations that have a sweep in this region.

### *Heat plot*

Heat maps for the whole turkey genome, including all turkey autosomes, and for the individual turkey chromosomes separately, were plotted to visualize overlapping signatures of selection in the different turkey populations using the "heatmap.plus" package in R [24]. The color scale is based on the square root of heterozygosity values, for visualization and distinction of sweep areas in the genomic regions.

### *Functional annotation analysis*

All genes lying within the overlapping sweep regions were used for functional annotation analysis. Functional annotation analysis was performed using DAVID (Database for Annotation, Visualization, and Integrated Discovery) with default parameters[25]. DAVID is a web-based bioinformatics application that systematically identifies enriched biology associated with large gene list(s) derived from high-throughput genomic experiments [25]. Correction for multiple comparison was done by the Benjamini-Hochberg method [26]. Annotation for turkey genes is very limited therefore we used one to one orthologous of turkey to human to perform this functional annotation analysis.

***Ethical approval for the use of animals in this study***

Although animals were used in this work, no direct experiments were performed on them. Blood sample collection was carried out by highly skilled and experienced personnel from the breeding companies. No approval from the ethics committee was necessary according to local legislation.

## 6.3 Results

In order to identify significant signatures of selection, threshold values were estimated for heterozygosity in each of the different turkey populations. These threshold heterozygosity values ranged from 1.0E-5 to 5.1E-5 (Table 6.1). The highest threshold value was obtained for L3 while the lowest threshold value was obtained for BvSW.

A whole genome view of the selective sweep regions in the different turkey populations is presented in figure 6.1. In total, we observed 54 significant genomic (P < 0.05) regions that were defined as signatures of selection on 14 different chromosomes in different turkey populations (Additional File 6.1). Areas with evidence of selective sweeps varied from 1.5 Mb to 11.1 Mb in length (Additional File 6.1).

Out of these 54 significant selective sweep regions, 31 were population specific (Additional File 6.1) and were observed on 12 different chromosomes, while 23 were overlapping selective sweep regions in multiple populations that were observed on 13 different chromosomes (Table 6.2 & Additional File 6.1 ). The majority of the population specific regions, 26 in total, were observed in the commercial populations (L1-L7), on average nearly 4 per population while heritage populations (BvSW, Nset and RP) showed 1.6 population specific sweep per population. Differences between commercial populations were considerable, with as many as 8 sweep regions observed in population L3 and only one population specific selective sweep region observed in population L6. Five population specific sweep regions were observed in heritage varieties with 1 (RP) or 2 (BvSW and Nset) sweeps per population.

Out of 23 sweep regions that showed overlap in multiple populations, one was observed only in the heritage varieties (Nset and RP) while 13 were observed only in the commercial lines (Table 6.2). Commercial line L1 had the largest sweep region, 11.1 Mb, (Additional File 6.1) as well as   the highest number (10) of overlapping selective sweep regions. The lowest number (3) of overlapping selective sweep regions was observed in the heritage variety Nset (Table 6.2).

**Table 6.1:** Estimates of threshold (≤0.05) of different turkey populations.

| ID | Threshold[1] |
|---|---|
| Commercial Line 1 | 0.000030 |
| Commercial Line 2 | 0.000022 |
| Commercial Line 3 | 0.000051 |
| Commercial Line 4 | 0.000014 |
| Commercial Line 5 | 0.000029 |
| Commercial Line 6 | 0.000032 |
| Commercial Line 7 | 0.000026 |
| Beltsville Small White | 0.000010 |
| Narragansett | 0.000049 |
| Royal Palm | 0.000023 |

[1] Estimates of threshold values (≤0.05) that determines probability of having number of heterozygous SNP count per nucleotide position in 5 consecutive bins of 300Kb for each population.

Differences were observed along the turkey genome, regarding the presence of sweeps at different chromosomes. Out of 54 observed sweep regions at different chromosomes, chromosome 2 showed the highest number of significant regions, 8 in total, while chromosome 14 showed the lowest number, 2 in total. Chromosomes 5, 7, 9 and 14 had five significant selective sweep regions that showed an overlap in at least 4 different populations (Table 6.2; Figure 6.1). Chromosome 5 had two overlapping selective sweep regions that were each shared by at least five populations, and one of these two regions was presented by commercial lines only (Table 6.2). Chromosome 9 also had a sweep region that was shared by five populations (Table 6.2 and Figure 6.1).

Overlapping selective sweep regions covered 5,452 genes, 34.7 % of the total number of genes that is identified in turkey genome sequence [17]. Out of these turkey genes, 3,858 were one to one orthologous with human genes and 3,832 turkey genes had a corresponding HGNC (HUGO Gene Nomenclature Committee) symbol in human genebuild. Finally, 3,718 of these genes with HGNC symbol had annotation information available in DAVID and were used in the functional annotation analysis. Functional annotation analyses resulted in 514 gene ontology (GO) terms with an Expression Analysis Systematic Explorer (EASE) P-value [27] of less than 0.1(Additional file 6.2) which is a rather liberal threshold because it does not correct for multiple testing. The EASE P-value is a modified Fisher Exact P-value. GO terms that passed the significant threshold of 0.05 after Benjamini Hochberg

correction [26] are shown in table 6.3. Several of the GO terms were found to be related with morphogenesis or growth (Additional file 6.2).
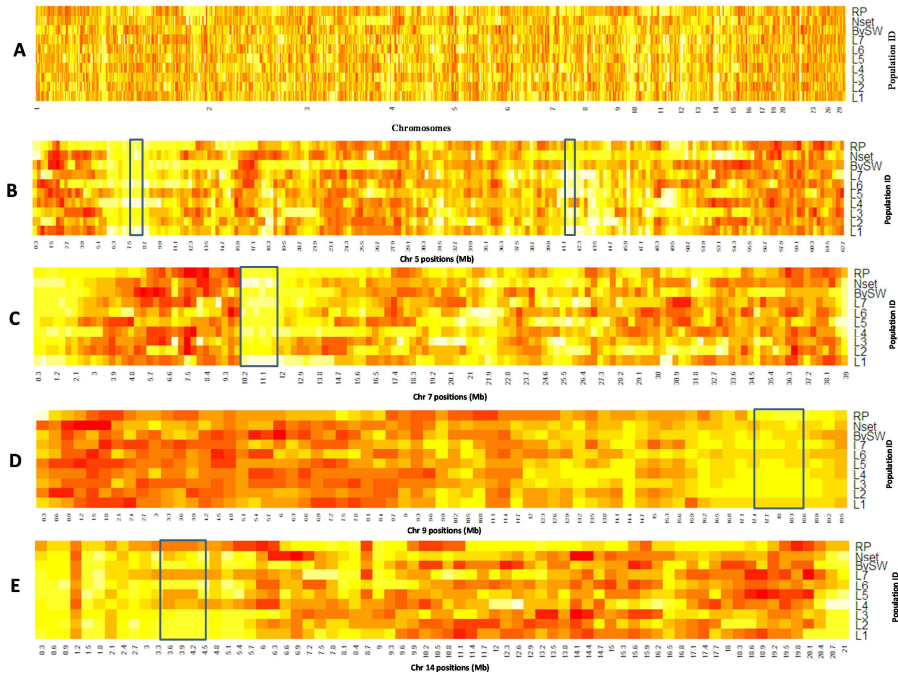


**Figure 6.1:** Genomic regions of selective sweep shared by different turkey populations. **A)** Turkey autosomes (1-30) showing variation in pattern of heterozygosity, colour pattern from light to dark shows low to high level of heterozygosity. **B)** Turkey chromosome 5 with 2 selective sweep regions from positions 7.8-8.4 Mb and 41.1-42 Mb shared by 5 different turkey populations L1, L4, L6, BvSW, RP and L1, L3, L5, L6, L7 respectively. **C)** Turkey chromosome 7 with selective sweep region from positions 9.9-11.7 Mb shared by 4 different turkey populations L1, L4, BvSW and RP. **D)** Turkey chromosome 9 with selective sweep region from positions 17.4-18.6 Mb shared by 5 different turkey populations L3, L5, L6, Nset and BvSW. **E)** Turkey chromosome 14 with selective sweep region from positions 3.3-4.5 Mb shared by 4 different turkey populations L1, L3, L6 and BvSW.

133

**Table 6.2:** Turkey selective sweeps showing overlap in multiple turkey populations.

| Chr | Sweep region[1] | Overlapping region[2] | Populations |
|---|---|---|---|
| 1 | 179100000-181800000 | 180000000-181800000 | Nset, RP |
| 2 | 48600000-51300000 | 49500000-51300000 | L5, L6 |
| 2 | 58200000-60300000 | 59700000-60300000 | BvSw, L3 |
| 2 | 83700000-85200000 | 83700000-85200000 | L1, L7 |
| 3 | 27600000-34200000 | 28800000-30300000 | L1, L5 |
| 3 | 35100000-37200000 | 35700000-37200000 | L1, L2, L4 |
| 3 | 96000000-97800000 | 96900000-97800000 | L1, L2, L5 |
| 3 | 96900000-99900000 | 98400000-99600000 | L2, L5, Nset |
| 4 | 48600000-51000000 | 49200000-51000000 | L2, L6 |
| 5 | 6900000-9000000 | 7800000-8400000 | L1, L4, L6, RP, BvSW |
| 5 | 41100000-42900000 | 41100000-42000000 | L1, L3, L5, L6, L7 |
| 6 | 7800000-9900000 | 7800000-9600000 | L1, L7 |
| 6 | 25200000-27900000 | 26700000-27900000 | L2, L3 |
| 7 | 9900000-12600000 | 9900000-11700000 | L1, L4, RP, BvSW |
| 8 | 300000-3300000 | 1200000-3000000 | L3, L5 |
| 9 | 12600000-14400000 | 13800000-14400000 | L4, L6 |
| 9 | 15900000-19500000 | 15600000-16200000 | L3, L5, L6 |
| 9 | 15900000-19500000 | 17400000-18600000 | L3, L5, L6, Nset, BvSW |
| 10 | 16800000-20100000 | 17400000-19200000 | L2, L5, RP |
| 11 | 1200000-8400000 | 4200000-7500000 | L2, L4, L7, RP |
| 11 | 7800000-12000000 | 9900000-12000000 | L3, L4 |
| 14 | 3000000-4500000 | 3300000-4500000 | L1, L3, L6, BvSW |
| 22 | 300000-2100000 | 600000-2100000 | L1, L3, L6 |

[1] Describes start and end positions of sweep regions at different turkey chromosomes. [2] Describes chromosome positions where overlap in sweep starts and ends in all populations that have sweep in this region.

**Table 6.3:** Gene ontology (GO) terms that passed significant threshold of 0.05 after Benjamini Hochberg correction.

| GO term | Annotation Term | Benjamini Hochberg P-value |
|---------|-----------------|---------------------------|
| GO:0048598 | Embryonic morphogenesis | 0.0005 |
| GO:0001882 | Nucleoside binding | 0.0022 |
| GO:0017076 | Purine nucleotide binding | 0.0029 |
| GO:0001883 | Purine nucleoside binding | 0.0043 |
| GO:0030554 | Adenyl nucleotide binding | 0.0045 |
| GO:0000166 | Nucleotide binding | 0.0080 |
| GO:0032553 | Ribonucleotide binding | 0.0091 |
| GO:0032555 | Purine ribonucleotide binding | 0.0091 |
| GO:0032559 | Adenyl ribonucleotide binding | 0.0155 |
| GO:0005524 | ATP binding | 0.0168 |

## 6.4 Discussion

We aimed at finding genomic regions with reduced heterozygosity, either resulting from strong selection in favor of specific alleles or from genetic drift. For the discovery of these regions in different turkey populations (commercial lines and the heritage varieties), we used a modified whole genome heterozygosity distribution approach [2]. In a particular population, the occurrence of heterozygosity values equal or less than the threshold value (Table 6.1) within at least 5 consecutive bins (each with 300 Kb size) indicates a significant reduction in heterozygosity in that region. In general, heterozygosity in turkey is low with an estimated average heterozygosity of 1.07 SNPs $Kb^{-1}$ [28], much lower than the observed heterozygosity in chicken, of 4.28 and 2.24 SNPs $Kb^{-1}$ reported in two different studies [2, 29]. We estimated threshold values separately for the different turkey populations. The threshold values (Table 6.1) can also be regarded as a measure of the level of genetic diversity in a particular population. In our study, we found the highest threshold value for commercial population L3, which is concordant with the highest observed genetic diversity and the highest number of SNPs discovered in this population in our previous study [28]. Similarly, the lowest threshold value was observed for BvSW, also concordant with the previously observed lowest genetic diversity and the lowest number of SNPs discovered in this population [28].

In our study, 48 significant regions (population specific and overlapping) were observed in the commercial populations while only 6 significant regions (population specific and overlapping) were observed in the heritage populations (Additional File

6.1 & Table 6.2). The presence of a high number of selective sweeps in commercial lines can be explained as a result of the high selection intensity applied to these populations [30]. A lower number of sweep regions in heritage varieties may be due to a number of reasons; admixture of populations, relatively high effective populations size in heritage varieties, or relatively less intensive and less specific directional selection applied to the heritage varieties in comparison to commercial turkeys. Specific information about population admixture or effective population size of heritage varieties is limited, but based on the anecdotal information from the turkey breeders, is expected to be low.

Regions with evidence for a selective sweep varied in size but generaly are very large (1.5-11.1 Mb). Reduction in genetic diversity and heterozygosity at different locations in the genome can persist for a long time, and indicate selection across a long genomic region [31]. The size of a sweep region may vary with history of domestication, the type of population (inbred or outbred), and intensity of selection within a particular population. SNP analyses of domestic dogs and cats show large stretches of alternating heterozygous and homozygous regions in both species as a consequence of domestication and breed development [32, 33]. Furthermore, in most outbred species, a selected region would display local SNP homozygosity, compared to abundant polymorphism elsewhere in the genome [34].

Uneven distribution of homozygous regions can be expected across the genome due to selection pressure through natural or artificial means [1-3, 35]. Chromosome 5, 7, 9 and 14 are highly distinct with overlapping regions in at least four different turkey populations (Table 6.2). This suggests that genomic region on these chromosomes contain gene(s) which affect the traits that are important for turkey production. Turkey populations that showed overlap in sweeps on these chromosomes may either be highly selected for specific objectives that all populations had in common or may have been developed from the common parents that already were homozygous for these sweep regions. Two significant selective sweep regions discovered on chromosome 5 and chromosome 22 show overlapping stretches only in commercial population (Additional File 6.1). These regions may contain genes involved in commercially important traits. These regions, however, are too large to identify the individual genes that may have been under selection.

Chromosome studies have revealed that the karyotype is much more conserved between avian species than in other taxa, such as e.g. in mammals, with most avian species showing a diploid chromosome number between 76 and 80 (http://www.genomesize.com). This shows that chromosomal evolution or large-

136

scale rearrangements affecting chromosome number occur at a low rate in birds, and as a result many chromosomes have remained more or less intact during avian evolution [36]. Comparative cytogenetic and linkage maps between turkey and chicken showed conserved synteny and close ancestral relation among these species [37, 38] and support the hypothetical ancestral Galliform karyotype [39]. Because of the strong structural as well as functional conservation between the turkey and the chicken [40, 41], as well as the similarities in breeding objectives, overlap in selective sweep regions may be expected. To test whether selective sweep regions are conserved between chicken and turkey, the orthology to chicken for all significant overlapping sweep regions of turkey was determined. These genomic regions were then examined for the presence of sweeps, based on two different studies on chicken [2, 42]. Selective sweep studies on chicken reported about 400 sweep regions [2, 42] which is about 0.38 sweep per Mb in chicken genome. Thirteen out of the 23 overlapping sweep regions identified in turkey, also harbored a selective sweep reported in chicken. Rubin et al. [2] reported 40 highly significant chicken sweep regions with very low Z transformed heterozygosity (ZHp < -6). Two of these highly significant chicken sweeps mapped within the syntenic regions of turkey sweeps on chromosomes 7 and 11 (Additional File 6.1). Overall the concordance of chicken sweep regions with turkey sweep regions was low. Approximately 0.32 chicken sweeps were observed per Mb within the total overlapping sweep length of turkey. This result shows no enrichment of chicken sweeps within the overlapping sweep regions of turkey.

The identified selective sweep regions are expected to have been involved in producing phenotypic variation for the traits of interest, which resulted in the fixation of these regions due to intensive selection. To investigate the variation explained by these regions, we looked for QTL information within these regions. Due to the limited availability of information on turkey QTLs and the presence of structural and functional conservation in the turkey and the chicken genomes [28, 38, 40], overlapping regions of significant selective sweeps (Table 6.2) of turkey were aligned with chicken genome sequence (WASHUC2) to determine their positions in the chicken genome (Additional file 6.3). The orthologous chicken regions were subsequently examined for the presence of reported chicken QTL for growth [43]. Many QTL were found to be overlapped with these genomic regions (Additional File 6.3). The frequency of chicken growth QTL for which the confidence interval overlapped with the turkey sweep regions was found to be 11.33 growth QTL per Mb of sweep region. This high frequency of chicken growth QTL overlapping with the turkey selective sweep regions was however a result of the

high number of growth QTL discovered in chicken. The sweep regions did not show an enrichment of chicken QTL compared to other parts of the genome.

Production census of turkeys from last four decades [44] show that turkeys have doubled in size. We had therefore expected to see a sweep in the region of (IGF-1), the somatomedin, insulin-like growth factor 1 which is well known to play an important role in muscle growth and development in various domesticated species [45-47]. We did not find a sweep near the IGF-1 region on turkey chromosome 1 (56348061bp-56402610bp). Previously, two QTL were detected in chicken for Insulin-like growth factor (IGF-1) levels in blood plasma, located at chromosome 1 and 2 [48, 49]. Both chicken QTL regions are syntenic with turkey and overlap with selective sweep regions at chromosome 1 and 6, respectively (Additional File 6.3). This suggests that specific genes present within the turkey sweep regions are involved in the insulin-like growth factor 1 hormone pathway, and that this pathway has also been under strong selection in turkey.

To understand the function of genes lying within the sweep regions of turkey, we performed a gene functional annotation analysis using DAVID. Our gene-based enrichment analysis showed enrichment of genes for regulation of development and morphogenesis within turkey sweep regions (Additional file 6.2). We found highly significant GO term with embryonic morphogenesis (Table 6.3) and other suggestive terms (Additional File 6.2) with e.g. embryonic organ morphogenesis, body development and maintenance of growth. This indicates that the observed sweep regions of turkey are enriched with genes that are important for growth and development.

## 6.5 Conclusion

The genome of commercial turkeys showed large selective sweep regions. The relatively high number of sweep regions in commercial populations in comparison to heritage varieties, and the enrichment of turkey sweep regions with genes of importance to growth, indicates that the turkey sweep regions are likely the result of intensive selection for growth moving specific haplotypes towards fixation.

## Authors' contributions

MLA and JWMB analysed the data. LAB assisted and trained laboratory personnel in genomic DNA isolation from blood and preparation of all libraries. CPVT assisted with project design and coordination. TSS helped develop sequencing strategy and supervised sequence production. SGS was responsible for sequencing processing. JAL organized blood sample collection and DNA sequencing, and was the Principal Investigator (Agriculture and Food Research Initiative Competitive Grant no. 2010-

## Additional files:

### Additional file 6.1
**Link:** http://vps6371.xlshosting.net/Aslam/
**Description:**
**Title: Position of turkey genomic regions with signatures of selection in different turkey populations.** This file contains the start and the end positions of turkey selective sweep regions on different chromosomes. This file also shows the positions for syntenic regions of turkey overlapping sweeps in chicken with reported chicken sweeps.

### Additional file 6.2
**Link:** http://vps6371.xlshosting.net/Aslam/
**Description:**
**Title: Gene ontology (GO) terms observed with functional annotation analysis performed using the DAVID.** This file contains all GO terms observed with functional annotation analysis with biological functions and P-values. This file also contains gene names that are involved in each and every GO term.

### Additional file 6.3
**Link:** http://vps6371.xlshosting.net/Aslam/

**Description:**

**Title: Syntenic positions of turkey overlapping sweep regions with chicken.** This file contains the start and the end positions of turkey overlapping selective sweep regions and their syntenic positions in chicken genome. It also contains information of discovered QTLs in chicken genome positions that are syntenic to turkey overlapping selective sweep regions.

# References

1.   Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: **A genome-wide scan for signatures of recent selection in Holstein cattle**. *Animal Genetics* 2010, **41**:377-389.

2.   Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al*: **Whole-genome resequencing reveals loci under selection during chicken domestication**. 2010, **464**(7288):587-591.

3.   Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, Thaller G, Simianer H: **Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle**. *BMC Genomics* 2011, **12**:318.

4.   Nielsen R, Bustamante C: **A scan for positively selected genes in the genomes of humans and chimpanzees**. *Plos Biology* 2005, **3**(6):e170.

5.   Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C *et al*: **Identification of ten loci associated with height highlights new biological pathways in human growth**. *Nat Genet* 2008, **40**(5):584-591.

6.   Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, Fall T, SeppÃ¤lÃ¤ EH, Hansen MST, Lawley CT *et al*: **Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping**. *PLoS Genet* 2011, **7**(10):e1002316.

7.   R. Howard, A. Moore: **A complete checklist of birds of the world**, Revised edn. London: Macmillan; 1984.

8.   Crawford RD: **Introduction to Europe and diffusion of domesticated turkeys from the America**. *Arch Zootec* 1992, **41** (extra):307-314.

9.   Hirst KK: **Turkey (*Melagris gallapavo* and spp.)**. In: *History of the Domestication of Turkeys.* vol. 2011: About.com Education Archaeology; 2011.

140

10.     Flannery KV: **Vertebrate fauna and hunting practices. Prehistory of the Tehuacan Valley**. In: *Environment and Subsistence.* Edited by Byers DS, vol. 1. Austin: Univ of Texas Press; 1967: 132-177.

11.     Schorger AW: **The Wild Turkey: Its History and Domestication**. In*.* Norman: Univ of Oklahoma Press; 1966.

12.     Association AP: **The American Standard of Perfection**. In*.* Troy, NY: APA; 2001: 492.

13.     SAS Institute: **SAS User's Guide: Statistics**. In*.*: SAS Institute Inc., Cary, NC; 2000.

14.     Maynard Smith J, Haigh J: **The hitchhiking effect of a favorable gene**. *Genet Res* 1974, **23**:23-35.

15.     Kijas JM, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S *et al*: **A genome wide survey of SNP variation reveals the genetic structure of sheep breeds**. *PLoS ONE* 2009, **4**(3):e4668.

16.     The SAM Format Specification Working Group: **The SAM format specification**. In., 1.4-r985 edn; 2011.

17.     Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Ann Blomberg L, Bouffard P, Burt DW, Crasta O, Crooijmans RPMA *et al*: **Multi-Platform Next-Generation Sequencing of the Domestic Turkey (*Meleagris gallopavo*): Genome Assembly and Analysis**. *PLoS Biol* 2010, **8**(9):e1000475.

18.     Kerstens HHD, Crooijmans RPMA, Veenendaal A, Dibbits BW, Chin-A-Woeng TFC, Dunnen JTd, Groenen MAM: **Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey**. *BMC Genomics* 2009, **10**:479.

19.     Li G, Ma L, Song C, Yang Z, Wang X, Huang H, Li Y, Li R, Zhang X, Yang H *et al*: **The YH database: the first Asian diploid genome database**. *Nucleic Acids Res* 2009, **37**:1025-1028.

20.     Marsden SJ: **The Beltsville small white turkey**. In*.* Beltsville Maryland, USA: Animal husbandry research division: 32-41.

21.     **The American Livestock Breeds Conservancy; Turkeys: Narragansett**

22.     **The American Livestock Breeds Conservancy; Turkeys: Royal Palm**

23.     Stromberg M: **Mosaik Assembler**. In: *110014.* Edited by Lee W-P: Boston College; 2010.

24.     Day A: **Heatmap with more sensible behavior**. In., 1.3 edn; 2007.

25.     Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. 2008, **4**(1):44-57.

26.     Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing**. *Journal of the Royal Statistical Society Series B-Methodological* 1995, **57**:289-300.

27.     Hosack D, Dennis G, Sherman B, Lane H, Lempicki R: **Identifying biological themes within lists of genes with EASE**. *Genome Biology* 2003, **4**(10):R70.

28.     Aslam ML, Bastiaansen JWM, Elferink M, Megens H-J, Crooijmans R, USDA, Long J, Groenen MAM: **Whole genome SNP discovery and the genetic diversity of different turkey populations**. *BMC Genomics* 2012:[Submitted].

29.     Wong GK-S, Liu B, Wang J, Zhang Y, Yang X, Zhang Z, Meng Q, Zhou J, Li D, Zhang J *et al*: **A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms**. *Nature* 2004, **432**(7018):717-722.

30.     McKay JC: **The genetics of modern commercial poultry**. In: *World's Poultry Congress: 2008; Brisbane, Australia,*; 2008: CD-ROM.

31.     The Chimpanzee Sequencing and Analysis Consortium: **Initial sequence of the chimpanzee genome and comparison with the human genome**. 2005, **437**(7055):69-87.

32.     Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC *et al*: **Genome sequence, comparative analysis and haplotype structure of the domestic dog**. 2005, **438**(7069):803-819.

33.     Pontius JU, Mullikin JC, Smith DR, Agencourt Sequencing Team, Lindblad-Toh K, Gnerre S, Clamp M, Chang J, Stephens R, Neelam B: **Initial sequence and comparative analysis of the cat genome**. *Genome Res* 2007, **17**:1675-1689.

34.     Oleksyk TK, Zhao K, De La Vega FM, Gilbert DA, O'Brien SJ, Smith MW: **Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations**. *PLoS One* 2008, **3**:e1712.

35.     Davidson S, Starkey A, MacKenzie A: **Evidence of uneven selective pressure on different subsets of the conserved human genome; implications for the significance of intronic and intergenic DNA**. *BMC Genomics* 2009, **10**:614.

36.     Nie W, O'Brien PCM, Ng BL, Fu B, Volobouev V, Carter NP, Ferguson-Smith MA, Yang F: **Avian comparative genomics: reciprocal chromosome**

painting between domestic chicken (Gallus gallus) and the stone curlew (Burhinus oedicnemus, Charadriiformes)—An atypical species with low diploid number. *Chromosome Res* 2009, **17**(1):99-113.

37.     Chaves LD, Rowe JA, Reed KM: **Survey of a cDNA library from the turkey (*Meleagris gallopavo*)**. *Genome* 2005, **48**:12-17.

38.     Griffin D, Robertson L, Tempest H, Vignal A, Fillon V, Crooijmans R, Groenen M, Deryusheva S, Gaginskaya E, Carre W *et al*: **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution**. *BMC Genomics* 2008, **9**(1):168.

39.     Griffin DK, Robertson LBW, Tempest HG, Skinner BM: **The evolution of the avian genome as revealed by molecular cytogenetics**. *Cytogenet Genome Res* 2007, **117**:64-77.

40.     Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Vereijken A, Megens H-J, Groenen MAM: **A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes**. *BMC Genomics* 2010, **11**:647.

41.     Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Vereijken A, Groenen MAM: **Whole genome QTL mapping for growth, meat quality and breast meat yield traits in turkey**. *BMC Genetics* 2011, **12:61**.

42.     Elferink MG, Megens H-J, Vereijken A, Hu X, Crooijmans RPMA, Groenen MAM: **Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds**. *PLoS One* 2012, **7**(2):e32720.

43.     Hu Z-L, Reecy JM: **Animal QTLdb: Beyond a Repository - A Public Platform for QTL Comparisons and Integration with Diverse Types of Structural Genomic Information**. *Mammalian Genome* 2007, **18**:1-4.

44.     **Food and agriculture organization statistical division (FAOSTAT) of the United Nations** [http://faostat.fao.org/]

45.     Klindt J, Yen JT, Buonomo FC, Roberts AJ, Wise T: **Growth, body composition, and endocrine responses to chronic administration of insulin-like growth factor I and(or) porcine growth hormone in pigs**. *J Anim Sci* 1998, **76**(9):2368-2381.

46.     Wooda BJ, Archerb JA, van der Werf JHJ: **Response to selection in beef cattle using *IGF-1* as a selection criterion for residual feed intake under different Australian breeding objectives**. *Liv Prod Sci* 2004, **91**(1-2):69-81.

47.     Sato S, Ohtake T, Uemoto Y, Okumura Y, Kobayashi E: **Polymorphism of insulin-like growth factor 1 gene is associated with breast muscle yields in chickens**. *Animal Science Journal* 2012, **83**(1):1-6.

48.     Park H-B, Jacobsson L, Wahlberg P, Siegel PB, Andersson L: **QTL analysis of body composition and metabolic traits in an intercross between chicken lines divergently selected for growth**. *Physiological Genomics* 2006, **25**(2):216-223.

49.     Nadaf J, Pitel F, Gilbert H, Duclos MJ, Vignoles F, Beaumont C, Vignal A, Porter TE, Cogburn LA, Aggrey SE *et al*: **QTL for several metabolic traits map to loci controlling growth and body composition in an F2 intercross between high- and low-growth chicken lines**. *Physiological Genomics* 2009, **38**(3):241-249.

50.     McMurtry JP, Francis GL, Upton FZ: **Insulinlike growth factors in poultry**. *Domest Anim Endocrinol* 1997, **14**:199-229.

# 7

## General Discussion

Understanding the etiology and biology of a trait is very important for the improvement of that trait. The genetic makeup is one of the factors that causes variation in the performance of individuals for specific traits of interest. Individuals can vary in their genetic makeup with respect to the specific base at a particular position in the genome (single nucleotide polymorphism, SNP) or with respect to structural variations (SVs) in the genome e.g. deletions, duplications, copy-number variants, insertions, inversions and translocations. The characterization of the genomic structure is one of the routes to provide insight into the genetic basis of a trait. Such genomic characterization of individuals can ultimately lead to the identification of the causative variants for a particular trait. The identification of the causative variant(s) that affect a particular phenotypic trait involves detection of genomic regions or markers that show evidence of association with that trait. Subsequently, variation within those genomic regions or marker genotypes are investigated to find the true causative variant(s) that can be used to move the trait in the desired direction. The identification of the genetic basis of a complex trait or disease is however not trivial. The availability of a high marker density in a genomic region will allow narrowing down the set of candidate mutations but the chance of pointing out the functional mutation is generally still difficult. In the last decade, with the availability of high-throughput sequencing technologies, many new tools such as reference genomes, large number of SNP markers and genome-wide assays with high marker densities have become available.

Currently it is not realistic for any species to characterize all its genomic variation because this would require the whole genome sequence and spatial location of every individual within a species [1] which is highly expensive. Without having the complete genome information at the level of all individuals, the development of genomic resources such as linkage maps and reference genome sequences is considered very important for a detailed understanding of traits (polygenic and monogenic) at the molecular level. Because a polygenic trait is influenced by multiple genes, the classical approach of mapping quantitative trait loci (QTL) using linkage maps can be helpful in the identification of genes/regions of the genome that have a major effect on the trait, but it is generally not very powerful for loci with a minor effect on the trait.

Below, I will discuss further considerations about results presented in this thesis, the state of genome research into the genetic control of complex traits, and future perspectives.

## 7.1 Genetic control of complex traits of economic importance

Estimating the heritability of a trait is a first step to know whether that trait can be improved by breeding and how fast it can be improved. Most of the economically important traits in livestock are polygenic traits e.g. body weight, egg production, egg size and meat tenderness. Many studies have reported the estimated heritabilities for economically important traits in different livestock species [2-6]. When a trait has a low heritability this indicates that non-additive genetic effects (dominance and epistasis) and/or the environment have a much larger influence on that trait than the additive genetic effects. An example of such a trait of low heritability is ultimate pH (pHu) as described in chapter 2 (heritability is 0.09). High heritability of a trait means that additive genetics plays a relatively large role in the variation of the trait phenotype, such as for body weight at 40 days (BW40) of age of turkey, which has a heritability of 0.42 (chapter 2). The heritability of a trait (low or high) will have an impact on selection decisions. Progress tends to be much slower in lowly heritable traits as for pHu but with a highly heritable trait such as BW40 rapid progress can be achieved due to greater accuracy in selection decisions [7]. Genetic markers are more important for traits with low heritabilities, traits that are difficult to measure or that can only be measured late in life and for sex specific traits.

Genetic correlations between traits are very important when multiple trait selection and improvement is considered. Genetic correlations are caused by two mechanisms; linkage (genes affecting two traits are located near to each other on the same chromosome and are transmitted from parent to offspring together) and pleiotropy (situation where one gene, or a group of genes, controls more than one trait). Pleiotropy is obviously a factor for the growth traits [8]. The genes affecting growth early in the growing period also affect growth later in life. In several studies growth traits have shown high positive genetic correlations [9-12]. Positive genetic correlation means that if you select to change one trait, a second trait moves in the same direction (both increase or both decrease). Likewise, genetic correlations may be negative, which means the traits respond in opposite directions (one increases when the other decreases). For example, studies have shown negative genetic correlations between growth and reproduction traits [5, 13-15] and we observed negative genetic correlations between percent drip loss (PDL) and ultimate pH (pHu) and between pHu and breast meat yield traits (Chapter 2). The biological explanation is that an increase in body weight, also results in an increase in body fat and in the production of a relatively high amount of lactic acid during rigor mortis which ultimately results in high drip loss. So, the improvement in the

quantity of meat may deteriorate the quality if selection is based solely on one of these antagonistically related traits. Marker assisted selection (MAS) can act as a beneficiary solution for such situations where selection is desired for antagonistically correlated traits [16]. To get to this beneficiary situation, however, the discovery of markers is needed that are affecting one trait without having unfavorable effects on the others. This situation may not come true, if QTL are not detected for, some of, the negatively correlated traits. For example, in our QTL study (Chapter 4), we did not find any significant QTL for pHu but we do find significant QTL for PDL even though these traits have high negative genetic correlation (Chapter 2).

The assessment of the total impact of selection must include correlations between the traits because changing one trait may also change performance of a correlated trait. On the other hand, correlations between traits may also be exploited to reduce testing costs. The use of correlated information can reduce costs and time if the target trait and the correlated trait are not having antagonistic relation or when the correlated trait is not in the breeding goal. The expected net effect of changing a trait will be the summation of the changes in the trait itself and all correlated traits. Breeders make selection decisions based on an index value which is obtained considering heritabilities of the traits, correlation among traits, and economic value of each trait [17]. Using this index value may improve one trait without harming others in case of negatively correlated traits.

A next step towards understanding genetic control of traits is to identify genomic regions that affect the performance for that particular trait. Such QTL mapping studies (chapter 4) require a genetic linkage map like the one described for turkey in chapter 3. For decades, the investigation of the genetic basis of complex economically important traits and diseases has been a major focus of scientists working in genetics of domesticated animals. The identification of QTL provides insight into important genetic questions such as the relative effect sizes and the number of regions or genes influencing a trait [18]. QTL mapping is used to address these questions but it requires knowledge of the phenotype and the pedigree, or the development of controlled crosses to generate a large number of progeny as well as genotyping of parents and offspring. Growth, breast yield and quality of meat is of interest to turkey breeders. The identification of QTL, markers and genes associated with meat quality would provide tools to improve meat quality traits. Very little is known about the genetic basis of economically important traits in turkey even though turkey is an intensively selected agriculturally important species. A probable reason of this limited emphasis on genetic research is the interest of turkey breeders for those traits (growth, breast yield and meat quality)

that can be improved at an acceptable pace without the use genomic information because most have moderate to high heritabilities. So, from a turkey breeder's perspective, knowledge of the genetic basis of the currently important turkey traits is more important to understand biology than to be used directly in selection.

When we compare the results presented in chapters 2 and 4 for meat quality traits drip and pHu, and relate this to the extensive literature on drip loss and pHu of meat, we see some unexpected QTL results. In our QTL study (chapter 4), we did not find any significant QTL for pHu. Given the high genetic correlation between PDL and pHu (chapter 2), physiological reasons (higher drip loss is related to higher lactic acid production which leads to low pH [19]), and the available evidence about the location of QTL for these traits in pigs [20], QTL for pHu were expected on at least a part of the chromosomes where QTL for PDL were detected. One explanation is that some regions on chromosomes 1, 4, 5 and 16 did show effects on both traits but the observed F-value for these region did not surpass the threshold for one or both traits (chapter 4). Regions showed an elevated F-value for pHu or PDL, but not for both traits with one exception on MGA5 (chapter 4 and figure 7.1). A region at MGA5 did show overlapping peaks for both traits but neither surpassed the significance threshold. This result could indicate that partially different sets of genes are involved in the control of these traits or, more likely, that there were differences in power to detect QTL for these traits.
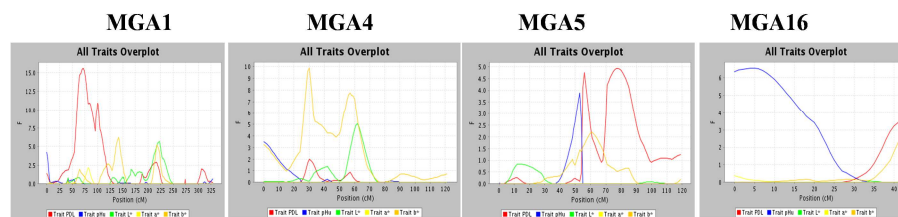


**Figure 7.1** Mapping of QTL at the turkey chromosomes 1, 4, 5 and 16. Percent Drip Loss(PDL), Ultimate pH (pHu), Lightness (L*), Redness (a*) and Yellowness (b*).

In our QTL study (chapter4), SNP variants were located within a number of growth related genes (*PIT1, AFABP, PRKAG3, IGF2* and *GDF8*) that were used to test whether these genes play a major role in the phenotypic variation. We did not find significant QTL near these genes but when these SNPs were included as fixed effects in the linkage analysis model, the F-value at the position of these SNPs decreased by more than 50% (Figure 7.2). The effect of these candidate SNPs on the traits, in combination with the absence of QTL in these regions may indicate that these SNPs are in linkage disequilibrium with a QTL in these regions which

segregates in both parent lines. This situation is contrary to the QTL mapping assumption that different QTL alleles are fixed in in the parent lines and explain why a QTL effect was not detected.
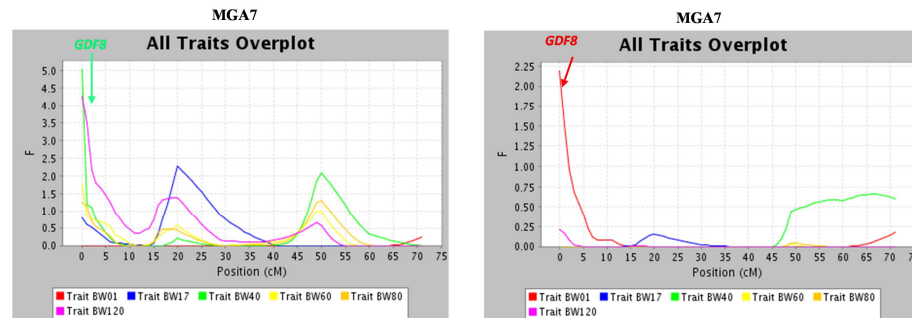


**Figure 7.2** Effect of SNPs selected from growth related gene *GDF8* and F-value reduces to half without the effect of SNPs. Day old body weight (BW01), body weight at day 17 (BW17), body weight at day 40 (BW40), body weight at day 60 (BW60), body weight at day 80 (BW80), and body weight at day 120 (BW120).

The results from our QTL mapping study indicate that to elucidate the genetic control of traits, need to perform more sophisticated experiments, such as analysis of extreme phenotype records (high vs. low performances), or experiments that include phenotypes across the whole scale of the trait, ideally on unrelated individuals, but more likely with a design that accounts for the effect of the pedigree.

The resolution of QTL linkage studies is low. Such studies often result in the identification of large genomic regions with hundreds of underlying genes. Increasing the number of animals or markers will only lead to limited improvements in the resolution of QTL linkage studies. QTL can be the result of a major gene linked to a specific phenotype or it may be the result of multiple variants within a cluster of linked genes each having a minor effect on the phenotype. The phase of the causative variants is important to be able to map a QTL resulting from multiple variants, which is automatically achieved in F2 analysis. Currently, the recommended strategy for the identification of a major gene(s) involves a genome wide association (GWA) analysis in combination with a search for functional candidate gene(s) and the discovery of variant/SNPs within these candidate gene. Several studies on domesticated animals have reported detection of causative variants in genes with major effect on phenotypes e.g. a missense variant in *DGAT1* affecting milk yield and milk composition in dairy cows[21], a

regulatory variant in *IGF2* affecting muscle growth in pigs [22] and a regulatory variant of *GDF8* affecting muscularity in sheep [23]. These studies describe genome-wide linkage analyses with a few hundred markers followed by fine-mapping and subsequently candidate gene sequencing to detect the real causative variant. In order to allow GWA analysis to find causative variants the density of SNP markers in the linkage map would not be sufficient (Chapter 3). Therefore we discovered millions of SNPs (Chapter 5) which will allow creation of high density SNP chip for further studies. As a next step towards the identification of causative variants, I would propose GWA studies with medium SNP densities, using around 60K SNPs, and genotyping outbred turkey populations of several thousand individuals.

GWA studies on human, have indicated that genetic variants detected for polygenic traits account for only a modest fraction (10–30%) of the observed heritability. This shows that even the utilization of high resolution SNP assays combined with large number of phenotyped individuals is not sufficient to detect all variants affecting complex polygenic traits. The proposed explanations for this missing heritability include a role of additional common variants of small effects [24], a role for less common variants [25], non-additive interactions between variants, and epigenetic effects [26]. Even with high density SNP assays, results from GWA studies suggest that we should know all genome-wide variants on a very large number of phenotyped individuals to detect all variants that affect the phenotype of polygenic traits. The identification of the causative variants that affect a trait will help in better understanding the biological processes and help to improve the traits in the desired direction. The identification of all variants within a population, currently is not realistic because of the relatively high cost of sequencing and phenotyping such a large number of individuals as well as analyzing such a large data set. In the near future, however, it will be possible to sequence a large number of individuals for a low price and then we may have performance and sequence data sets of many individuals. The handling of these huge data sets will require modified or novel smart methods and tools to get biological meanings out of them.

## 7.2 Turkey Breeding
### *Domestication*
Animal domestication revolutionized the lives of human in the past millenia. There is very little information available about the domestication history of turkey compared to other domesticated livestock species. Different schools of thoughts exist about the domestication of turkey (Chapter 1). Phylogeographic analyses described by Speller et al. [27] indicated that domestic turkey descends from the

South Mexican (SM) turkey (*Meleagris gallopavo gallopavo)* indigenous to Mexico [27]. In our study (chapter 5), we used SM individuals from a wild population to root the phylogenetic tree. No sequence data was available of the turkey subspecies from North America. The availability of sequences from a higher number of individuals per population/subspecies will allow the development of population specific haplotypes and the analysis of haplotype sharing among the populations. This could lead to identifying the involvement of the different subspecies in the development of modern commercial turkeys. Haplotype based analyses are considered better than SNPs based analyses in finding relationships among populations due to the higher specificity of haplotypes. Knowledge about the genomic similarities could guide the management or introduction of genetic diversity in inbred populations. We know for different species that today's domesticated populations have lower genetic diversity than the wild populations and we observe the same in turkey. For the breeding industry there is no need for introgression because genetic improvement is still achieved and the level of inbreeding is being controlled.

Heritability of economically important traits (growth, meat quality) in turkey are comparable with the heritabilities of those traits in chicken[19, 28]. Even though the genetic diversity in turkey is lower compared to chicken, the heritabilities are comparable and genetic improvement is being obtained, which indicates that diversity in turkey at the loci that contribute to genetic variation is still considerable.

### *Turkey genetic diversity*

Uncovering the genetic relationships between the commercial and the heritage populations of turkey reveals the genetic diversity available to breeders. Genetic diversity is the sequence variation within the species. Information on genetic diversity and relationships among and between individuals, populations, breeds, species, varieties, is of high importance to breeders for genetic improvement, conservation biology and for studying the evolutionary ecology of populations. We used the identified SNPs (chapter 5) to estimate relatedness among the sequenced turkey populations. Information of genetic diversity can be used in the design of breeding programs including making decisions on introgression of novel genes that may affect economically important traits such as growth, meat quality, fitness, and survival traits [29, 30].

Low SNP density (Chapter 5) can be a result of selection pressure and the domestication history of turkey. Continuous selection in the same direction can lead to complete loss of certain alleles from the domesticated populations and as a result a decreased genetic diversity of the population [31-33]. Muir et al. [34]

showed in chicken that commercial lines (broiler and layers) are missing a significant fraction of the genetic diversity found in non-commercial chickens. In chapter 5, we also observed a lower genetic variation (a relatively lower heterozygosity) in commercial turkey populations compared to the heritage and the wild turkey populations (Chapter 5) along with a higher number of large stretches of homozygosity in commercial turkey lines (Chapter 6). Even with the limited sequence data for wild turkey and about 6 times more data for commercial populations, we still observed different alleles in wild turkey that were (nearly) lost in current turkeys. The presence of different alleles in the wild SM and the domesticated populations is a demonstration of their divergence during the course of domestication (Chapter 5). Our results indicate that domesticated turkey lines have been selected (artificially or naturally) for non-wild-type alleles. (Chapter 5).

In a more diverse gene pool, breeders could select for alleles that might be related to resistance for a specific disease, affect the ability of the organism to survive under different environmental conditions, or that might be related to certain traits related to new breeding objectives. Within a more narrow gene pool, breeders may not be able to find alleles to improve a trait in the desired direction. In animal breeding, crosses with non-commercial populations are rarely applied. Rather, genetically improved animals are often kept in small, closed populations. This intense selection leads to reduced polymorphism levels and increased LD in domesticated species [34-37]. The commercial turkey industry is dominated by only two multinational breeding companies (Aviagen and Hybrid). The consolidation of the industry will have led to a reduction in the number of commercial populations which, in addition to intensive selection in the commercial lines, may also have contributed to the reduced genetic diversity of turkey. However, a similar consolidation is seen in the chicken breeding industry and there the genetic diversity is higher.

The turkey is registered as a single breed [38, 39] with eight different varieties that are defined primarily by plumage colour (Bronze, Narragansett, White Holland, Black and Slate, Beltsville Small White, Bourbon Red, and the Royal Palm, registered in 1971[40, 41]). Inclusion of different turkey varieties in a single breed already provides an indication of the presence of lower variation in turkey. In our study (Chapter 5), Nei's genetic distances between the individuals from different turkey varieties (heritage breeds) were found to be in the range of 0.05-0.14 (results not shown). Nset2 and BvSW1 showed the lowest genetic distance while BvSW3 and RP1 showed the highest genetic distance (0.14). We compared Nei's genetic distances from our study on turkey with the Nei's genetic distances among different breeds of chicken and found that some chicken breeds (White Leghorn Vs.

Rhode Island Red; White-Faced Spanish Vs. Black Menorca) have lower Nei's genetic distance [42, 43] than some of the different turkey varieties. So, based on genetic distances we could consider turkey varieties as different breeds.

The SNP rate *within* turkey varieties is much smaller than SNP rate within chicken breeds. However, genetic diversity *between* turkey varieties is comparable to genetic diversity between chicken breeds. This low genome variation within, and high genetic diversity between turkey varieties suggests that sharing of alleles between the different populations is more limited in turkey than in chicken. Given the presence of more unique alleles in different populations, genetically diverse turkey lines could be created by introgression of genes or by selection across populations. Such genetically diverse turkey populations have probably been lost due to intensive selection or drift within small populations.

*Commercial breeding*

Today, only two international companies own most of the commercial turkey genetic resources, Aviagen, based in the United Kingdom, and Hybrid, based in Canada. Currently, selection in turkeys is based on multi trait BLUP (best linear unbiased prediction) breeding values including all information (production and reproduction traits). There is also some culling on individual phenotypic information. At present, no genomic information (QTL information) is used in selection probably because of different reasons: high genotyping cost, unavailability of high density (HD) SNP chip for turkey, reduced competition between breeders, and also the interest of breeders in traits (growth and meat quality traits) that have medium to high heritability and are easy to measure. This means that all QTLs based on microsatellites or low density SNP chip are not used. The presence of more turkey breeders might increase the chance that one of them will start to apply advanced methods of selection (including genomic information) like the methods that are used in some other livestock species [44, 45]. In the future, if one breeder starts to apply these tools, then it is very likely that others will also follow this approach. Turkey breeding is different from chicken breeding with respect to the selection objectives. In chicken, breeders are applying genomic selection for the improvement of traits [44] because a wider range of traits is included in the selection objectives such as reproductive and health related traits. Nevertheless, genomic information may be applied and can be helpful in the improvement of e.g. reproduction related traits in turkey.

The studies described in chapters 5 and 6 of this thesis included seven commercial lines that were used to analyse genetic diversity and discover stretches of homozygosity in the turkey genome. We obtained a large number of SNP markers that may be used by turkey breeders for future developments in turkey breeding.

155

For example in follow up studies using these millions of discovered SNPs (chapter 5), a high-density SNP chip can be developed for performing GWA studies to discover variant (s) for both minor and major effects on traits of economic importance, and a high-resolution linkage map can be developed. The knowledge of the biological mechanisms how a particular variant affects a specific trait, is important for the further improvement of the trait in the desired direction.

A novel development to use markers, which has found rapid uptake in other species, is called genomic selection. Genomic Selection allows for direct selection between individuals that don't have own performance on sex-limited traits (e.g. egg production or eggshell quality in males), sex-influenced traits (e.g. body weight), traits that are difficult or expensive to measure (e.g. disease resistance, feed efficiency), or traits that require a long time to get the information (e.g. persistency, advanced age performance traits). Genomic selection is the newest tool available to the poultry breeding industry for genetic improvement [44] but looking beyond genomic selection we may see a renewed role for marker assisted selection. The availability of high a density SNP chip, high density genotypic information on parents, imputation of progeny genotypes from their own low density genotypes should help in reduction of phasing problem that was a limiting factor for breeders to apply marker assisted selection.

### *Challenges*

Integrating genomic information into existing breeding programs is one of the biggest challenges in (turkey) breeding. Current traits in the breeding program are those for which sufficient information can be obtained easily while the use of genomic information is limited and still relatively expensive. There are, however, other traits which are expensive and difficult to measure for which the use of genomic information will be cost effective. Genotyping cost of individuals is still the limiting factor. Cost per SNP genotype is declining but per sample cost is still high. With the rapid advancement in next generation sequencing (NGS) technologies the cost of re-sequencing is also decreasing. In the near future, the use of next generation sequencing at low cost will make it possible for breeders to sequence all individuals and perform association studies including total genomic variation present in the population.

The accuracy of predicting breeding values as well as the rate of genetic improvement can be maximized by combining genomic as well as trait information in an optimal way. Selection of individuals can be done using only genomic information without the need for trait measurements which saves time and effort in the measurement of trait. I agree that this is true but the genomic information related to the trait of interest needs to be developed using reference populations

156

with phenotypic records. This will be the reality for turkey within a few years. Novel statistical approaches, including single-step whole genome selection, are under development and will provide the framework for integration of phenotypic, pedigree, and genomic information for selection [44].

The turkey genome assembly is incomplete and, like the chicken genome assembly, it is still missing the sequence for most of the smallest microchromosomes. Sequencing of these microchromosomes is a big challenge because they are high in GC content. Sequencing of these microchromosomes is very important because of presence of high gene density (chapter 3 [46]). The lack of sequence information on microchromosomes and some of the other regions of the genome is a hurdle for genomic research in turkey. Genes or variants that are located on these microchromosomes or regions that effect phenotypic variance cannot be detected until their sequence information is available.

## References

1.    Gillespie JH: **Population genetics**. In: *JHU Press.* 2004.

2.    Arthur JA, Abplanalp H: **Linear estimates of heritability and genetic correlation for egg production, body weight, conformation and egg weight of turkeys**. *Poult Sci* 1975, **54**:11-23.

3.    Grossman M, Bohren BB: **Logistic growth curve of chickens: heritability of parameters**. *J Hered* 1985, **76**:459-462.

4.    Koerhuis ANM: **Non-normality of egg production distributions in poultry and the effects of outlier elimination and transformation on size and curvilinearity of heritability**. *Livest Prod Sci* 1996, **45**:69-85.

5.    Nestor KE, Anderson JW, Patterson RA: **Genetics of growth and reproduction in the turkey. 14. Changes in genetic parameters over thirty generations of selection for increased body weight**. *Poult Sci* 2000, **79**(4):445-452.

6.    Rodenburg TB, Buitenhuis AJ, Ask B, Uitdehaag KA, Koene P, van der Poel JJ, Bovenhuis H: **Heritability of feather pecking and open-field response of laying hens at two different ages**. *Poult Sci* 2003, **82**:861-867.

7.    Cassell B: **Using heritability for genetic improvement**. In*.* Virginia: Communications and Marketing, College of Agriculture and Life Sciences, Virginia Polytechnic Institute and State University; 2009.

8.    Nagamine Y, Pong-Wong R, Visscher PM, Haley CS: **Detection of multiple quantitative trait loci and their pleiotropic effects in outbred pig populations**. *Genetics selection evolution* 2009, **41**:44.

9.      Chapuis H, Tixer-Boichard M, Delabrosse Y, Ducroc V: **Multivariate restricted maximum likelihood estimation of genetic parameters for production traits in three selected turkey strains**. *Genet Sel Evol* 1996, **28**:299–317.

10.     Myers JM, Hershberger WK, Saxton AM, Iwamoto RN: **Estimates of genetic and phenotypic parameters for length and weight of marine net-pen reared coho salmon (Oncorhynchus kisutch Walbaum)**. *Aquac Res* 2001b, **32**:277–285.

11.     Neira R, Lhorente JP, Araneda C, Diaz NF, Bustos E, Alert A: **Studies on carcass quality traits in two populations of coho salmon (Oncorhynchus kisutch): phenotypic and genetic parameters**. *Aquaculture* 2004, **231**:117-131.

12.     Kranis A, Hocking PM, Hill WG, Woolliams JA: **Genetic parameters for a heavy female turkey line: impact of simultaneous selection for body weight and total egg number**. *Br Poult Sci* 2006, **47**(6):685 - 693.

13.     Nestor KE, McCartney MG, Harvey WR: **Genetics of growth and reproduction in the turkey. 1. Genetic and non-genetic variation in body weight and body measurements**. *Poult Sci* 1967, **46**:1374-1384.

14.     Nestor KE, Anderson JW, Patterson RA, Velleman SG: **Genetics of growth and reproduction in the turkey. 17. changes in genetic parameters over forty generations of selection for increased sixteen-week body weight**. *Poult Sci* 2008, **87**:1971-1979.

15.     Dana N, Waaij EHV, van Arendonk JAM: **Genetic and phenotypic parameter estimates for body weights and egg production in Horro chicken of Ethiopia**. *Trop Anim Health Prod* 2011, **43**(1):21-28.

16.     Verhaegen D, Plomion C, Gion J-M, Poitel M, Costa P, Kremer A: **Quantitative trait dissection analysis in Eucalyptus using RAPD markers: 1.Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages**. *Theor Appl Genet,* 1997, **95**:597-608.

17.     Stewart TS, Neal SM, Irvin KM: **Multiple trait selection for pork improvement: Swine genetics**. In.: Purdue University, Ohio State University Agricultural Technical Institute, Ohio State University,; 1999.

18.     Orr HA: **The population genetics of adaptation: the distribution of factors fixed during adaptive evolution**. *Evolution* 1998, **52**:935-949.

19.     Le Bihan-Duval E, Debut M, Berri CM, Sellier N, Santé-Lhoutellier V, Jégo Y, Beaumont C: **Chicken meat quality : genetic variability and relationship with growth and muscle characteristics**. *BMC Genetics* 2008, **9**:53.

158

20. Li HD, Lund MS, Christensen OF, Gregersen VR, Henckel P, Bendixen C: **Quantitative trait loci analysis of swine meat quality traits**. *Journal of Animal Science* 2010, **88**(9):2904-2912.

21. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P *et al*: **Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition**. *Genome Res* 2002, **12**:222-231.

22. Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M *et al*: **A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig**. *Nature* 2003, **425**:832-836.

23. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen J-M, Eychenne F *et al*: **A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep**. *Nat Genet* 2006, **38**(7):813-818.

24. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW *et al*: **Common SNPs explain a large proportion of the heritability for human height**. 2010, **42**(7):565-569.

25. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al*: **Finding the missing heritability of complex diseases**. 2009, **461**(7265):747-753.

26. McCarthy MI, Hirschhorn JN: **Genome-wide association studies: potential next steps on a genetic journey**. *Human Molecular Genetics* 2008, **17**(R2):R156-R165.

27. Speller CF, Kemp BM, Wyatt SD, Monroe C, Lipe WD, Arndt UM, Yang DY: **Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication**. *PNAS* 2010, **107**(7):2807-2812

28. Adeyinka IA, Oni OO, Nwagu BI, Adeyinka FD: **Genetic Parameter Estimates of Body Weights of Naked Neck Broiler Chickens**. *Int J Poult Sci* 2006, **5**(6):589-592.

29. Notter DR: **The importance of genetic diversity in livestock populations of the future**. *J Anim Sci* 1999, **77**:61-69.

30. Bijma P, Meuwissen THE, Woolliams JA: **Design of sustainable breeding programs in developed countries**. In: *7th World Congress on Genetics Applied to Livestock Production: 2002*; 2002: CDROM paper 24–01.

31. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: **A genome-wide scan for signatures of recent selection in Holstein cattle**. *Animal Genetics* 2010, **41**:377-389.

32. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al*: **Whole-genome resequencing reveals loci under selection during chicken domestication**. 2010, **464**(7288):587-591.

33. Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MAM: **Genome-Wide Footprints of Pig Domestication and Selection Revealed through Massive Parallel Sequencing of Pooled DNA**. *PLoS One* 2011, **6**(4):e14782.

34. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, Megens H-J, Zhang H, Okimoto R, Vereijken A *et al*: **Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds**. *PNAS* 2008, **105**(45):17312-17317.

35. Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M: **Extensive genome-wide linkage disequilibrium in cattle**. *Genome Res* 2000, **10**:220-227.

36. McRae AF, McEwan JC, Dodds KJ, Wilson T, Crawford AM, Slate J: **Linkage disequilibrium in domestic sheep**. *Genetics* 2002, **160**(3):1113-1122.

37. Amaral AJ, Megens H-J, Crooijmans RPMA, Heuven HCM, Groenen MAM: **Linkage disequilibrium decay and haplotype block structure in the Pig**. *Genetics* 2008, **179**(1):569-579.

38. Jacob J, Pescatore T, Cantor A: **Selecting the right turkey variety**. In*.* Lexington, UK: University of Kentucky College of Agriculture.

39. Austic RE, Nesheim MC: **Poultry Production**. In*.*, 13 edn. Lea and Febiger, Philadelphia; 1990.

40. Frank R, Reese J, Bender M, Sponenberg DP, Williamson D, Beranger J: **Selecting your best turkeys for breeding**. In*.*: The American Livestock Breeds Conservancy.

41. American Poultry Association: **APA recognized breeds and varieties**. In*.*; 2010.

42. Dávila SG, Gil MG, Resino-Talaván P, Campo JL: **Evaluation of diversity between different Spanish chicken breeds, a tester line, and a White Leghorn population based on microsatellite markers**. *Poult Sci* 2009, **88**:2518-2525.

160

43.     Mollah MBR, Islam FB, Islam MS, Ali MA, Alam MS: **Analysis of genetic diversity in Bangladeshi chicken using RAPD markers**. *Biotechnology* 2009, **8**(4):462-467.

44.     Fulton JE: **Genomic selection for poultry breeding**. *Animal Frontiers* 2012, **2**(1):30-36.

45.     Schefers JM, Weigel KA: **Genomic selection in dairy cattle: Integration of DNA testing into breeding programs**. *Animal Frontiers* 2012, **2**(1):4-9.

46.     International Chicken Polymorphism Map Consortium: **A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms.** *Nature* 2004, **432**:717-722.

# Summary

## Summary

The turkey (*Meleagris gallopavo*, MGA) is an important agricultural species that is largely used as a meat type bird as egg production of this species is very low. Turkey is the second largest contributor to the world's poultry meat production after chicken. Understanding the etiology and biology underlying production or health traits is very important for the genetic improvement of these traits in the desired direction and to avoid undesired side-effects. The aim of the research described in this thesis was to explore the genetics related to turkey production and to investigate genomics of turkey.

Chapter 1 of this thesis provides an overview of the domestication history of the different varieties of turkey, the genetic nature of different traits, and the importance to identify genetic variants affecting these traits. Different genetic tools and techniques are also discussed that either lead to the detection of genomic regions that affect different traits of economic importance. or to resources that allow the identification of genetic variation.

We had access to a turkey population that was based on parents from two lines that were crossed to produce full-sib families in the F1 generation. An F2 generation of 18 full sib families was produced by crossing 17 randomly selected F1 males and 18 randomly selected F1 females. Several phenotypic traits were measured and recorded on individuals of the F2 generation (1,716 offspring). Chapter 2, is about the estimation of genetic and (common) environmental variances for different growth (body weight and growth curve traits), breast meat yield and meat quality traits in turkeys. This study provides heritabilities, genetic and phenotypic correlations between the different traits which are important when considering multiple trait improvement and selection. This study showed that body weight traits, upper asymptote (a growth curve trait), percent breast meat and redness of meat had high heritabilities whereas heritabilities of breast length, breast width, percent drip loss, ultimate pH, lightness and yellowness of meat were medium to low. High positive unfavourable genetic and phenotypic correlations were found between body weight, upper asymptote, most breast meat yield traits and percent drip loss. Percent drip loss was found strongly negatively correlated with ultimate pH. The results of this analysis and the growth curve from the studied population of turkey birds suggest that in turkey, birds could be selected for breeding between 60 and 80 days of age in order to improve overall production and the production of desirable cuts of meat. The continuous selection of birds within this age range could promote high growth rates but specific attention to meat quality would be needed to avoid its deterioration due to the presence of antagonistic genetic correlations between meat quantity and quality.

**Summary**

In chapter 3, we describe the construction of a single nucleotide polymorphism (SNP) based linkage map of turkey and its comparison with the physical map of chicken to investigate genome structural differences between these highly important poultry species. In this study, eighteen full sib families, comprising 1008 (35 F1 and 973 F2) birds, were genotyped for 775 SNPs. Of the 775 SNPs, 570 were informative and used to construct a linkage map in turkey. The final map contained 531 markers in 28 linkage groups. The total genetic distance covered by these linkage groups was 2,324 centimorgans (cM) with the largest linkage group (81 loci) measuring 326 cM. Average marker interval for all markers across the 28 linkage groups was 4.6 cM. Comparative mapping of turkey and chicken confirmed two inter-, and 57 intra-chromosomal rearrangements between these two species which lead to the conclusion that turkey and chicken have highly conserved genomic structure with a relatively low number of inter-, and intra-chromosomal rearrangements.

In chapter 4, we used the linkage map of turkey that was developed in chapter 3 of this thesis, together with individual phenotypes to map quantitative trait loci (QTL) in the same population for different growth curve, body weight, breast yield and the meat quality traits. Results showed QTL on 21 of the 27 turkey chromosomes covered by the linkage map. Forty-five quantitative trait loci were detected across all traits and these were found in 29 different regions on the 21 chromosomes. Out of the 45 QTL, twelve showed significant (chromosome wide $P < 0.01$) evidence of linkage while the remaining 33 showed suggestive evidence (chromosome wide $P < 0.05$) of linkage with different growth, growth curve, meat quality and breast yield traits. In this study pleiotropic effects or close linkages between QTL were suggested for several of the chromosomal regions. The comparative analysis regarding the location of QTL discovered in turkey, and the reported chicken QTL (QTLdb) on the syntenic chicken chromosomes for the same, or a very similar trait, revealed signs of functional conservation between these species.

The next step after the use of the reference population in exploring genetics related to turkey production (estimation of genetic parameters, construction of linkage map and the QTL map) was to investigate the genomics of turkey variation (SNPs discovery, population diversity and signatures of selection). Chapters 5 and 6 describe the use of next generation sequencing to investigate genome variation and the discovery of genome-wide signatures of selection in the turkey respectively. Next generation sequencing was performed on 32 individuals from eleven different turkey populations (seven commercial, three heritage and a South Mexican wild population). Chapter 5 describes whole genome SNP discovery in turkey that resulted in the detection of 5.49 million putative SNPs compared to the

reference genome. The SNPs discovered were subsequently used for the analysis of genetic diversity between the different populations. A total of 75,254 SNPs were discovered in exonic regions, consisting of 23,795 non-synonymous, 52,506 synonymous, 377 stop gain and 8 stop loss variants. The average frequency of heterozygous nucleotide positions in individual turkeys was 1.07 Kb$^{-1}$. This low level of heterozygity in turkey genome relative to other livestock species like chicken and pig indicates much less genomic diversity in the turkey genome. The occurrence of low heterozygosity among commercial lines, as well as the presence of alleles/haplotypes in the wild South Mexican population that were not found in domesticated populations, underscores that specific haplotypes have been selected for, or have been lost in the modern domesticated turkey. In a genetic diversity analysis, all of the commercial lines branched from a single node relative to the heritage varieties and the ancestral turkey population, indicating that commercial lines appear to share a common origin.

In chapter 6, the next generation sequencing data from ten of the turkey populations (29 individuals) was used to detect selective sweep regions. Across the turkey populations we observed 54 genomic regions with significant (P < 0.05) evidence for a selective sweep. These sweeps were distributed over 14 different chromosomes. Out of these 54 significant selective sweep regions, 31 were population specific while 23 showed overlap with a selective sweep region in one or more populations. The 23 overlapping selective sweep regions were distributed over 13 different turkey chromosomes. Out of the 31 population specific selective sweep regions, 26 were found in the commercial populations. The size of the observed selective sweep regions was large. The relatively high number of selective sweep regions in commercial turkey populations, in comparison to turkey heritage varieties, and the enrichment and the enrichment of turkey sweep regions with genes of importance to growth, indicates that the selective sweep regions in turkey are likely the result of intensive selection for growth, moving specific haplotypes towards fixation.

Finally, in chapter 7, I discuss the main findings of this thesis with respect to their implication for breeding and selection. In this chapter the roles of genes and the genome in controlling complex economically important traits is discussed. In this chapter, I also discuss ways/techniques to identify variants in the genome that affect the performance of individuals for specific traits. The effect of domestication and selection on turkey diversity is described. Furthermore, I describe current selection procedures applied in the turkey industry, challenges faced by the breeders in application of genomic selection and how the use of genomic information can be effective in turkey industry.

# Samenvatting

# Samenvatting

De kalkoen (*Meleagris gallopavo*, MGA) is een belangrijk landbouwhuisdier dat vooral wordt gebruikt voor vleesproductie omdat de eierproductie van kalkoenen zeer laag is. Na kip is kalkoen het meest geproduceerde pluimvee vlees in de wereld. Inzicht in de etiologie en de biologie die ten grondslag liggen aan productie- of gezondheidskenmerken is zeer belangrijk voor de genetische verbetering van deze kenmerken in de gewenste richting en om ongewenste neveneffecten te voorkomen. Het doel van het onderzoek beschreven in dit proefschrift was om de genetica die verband houden met kalkoen productie te verkennen en het genoom van kalkoen te onderzoeken.

**Hoofdstuk 1** van dit proefschrift geeft een overzicht van de domesticatie geschiedenis van de verschillende kalkoenrassen, de genetische aard van verschillende kenmerken, en het belang om genetische varianten te identificeren die een effect hebben op deze kenmerken. Verschillende genetische methodes en technieken, die ofwel leiden tot het opsporen van regio's in het genoom die economisch belangrijke kenmerken beïnvloeden, of leiden tot de identificatie van genetische variatie, worden bediscussieerd.

Een kalkoen populatie gebaseerd op ouders van twee lijnen die gekruist zijn om full-sib families te vormen in de F1-generatie was beschikbaar voor deze studie. Een F2-generatie van 18 full-sib families werd geproduceerd door het kruisen van 17 willekeurig geselecteerde F1 mannelijke dieren en 18 willekeurig gekozen F1 vrouwelijke dieren. Verschillende fenotypische kenmerken werden gemeten aan individuen van de F2-generatie (1.716 nakomelingen). **Hoofdstuk 2** gaat over het schatten van genetische variantie en (gemeenschappelijk) milieu variantie voor verschillende groei kenmerken (lichaamsgewicht en groeicurve), borstvlees opbrengst en vleeskwaliteit kenmerken bij kalkoenen. Deze studie resulteerde in erfelijkheidsgraden, genetische en fenotypische correlaties tussen de verschillende kenmerken die van belang zijn bij genetische verbetering en selectie gebaseerd op meerdere kenmerken. Deze studie toonde aan dat lichaamsgewicht kenmerken, bovenste asymptoot (een groeicurve kenmerk), percentage borstvlees en roodheid van vlees hoge erfelijkheidsgraden hadden, terwijl de erfelijkheidsgraden van borstlengte, borstbreedte, percentage drip, pH van het vlees, lichtheid en gele verkleuring van vlees gemiddeld tot laag waren. Hoge positieve maar ongunstige genetische en fenotypische correlaties waren gevonden tussen lichaamsgewicht, bovenste asymptoot, de meeste borstvlees kenmerken en het percentage drip. Percentage drip bleek sterk negatief gecorreleerd met pH van het vlees. De resultaten van deze analyse en de groeicurve van de bestudeerde populatie kalkoenen suggereren dat kalkoenen op een leeftijd van 60 tot 80 dagen oud

kunnen worden geselecteerd voor de fokkerij, om de totale productie en de productie van gewenste stukken vlees te verbeteren. Continue selectie van vogels in deze leeftijdscategorie kan de groeicijfers bevorderen, maar dan is er specifieke aandacht nodig voor de kwaliteit van het vlees om de verslechtering ervan, door de aanwezigheid van antagonistische genetische correlaties tussen vlees kwantiteit en kwaliteit, te voorkomen.

In **hoofdstuk 3** beschrijven we de constructie van een genetische kaart voor kalkoen op basis van single nucleotide polymorfisme (SNP) en vergelijken die met de fysieke kaart van de kip om structurele verschillen in het genoom van deze zeer belangrijke pluimveesoorten te onderzoeken. In dit onderzoek werden achttien full-sib families, bestaande uit 1.008 (35 F1 en F2 973) vogels, gegenotypeerd voor 775 SNPs. Van de 775 SNPs waren er 570 informatief en gebruikt om een genetische kaart voor kalkoen te bouwen. De uiteindelijke genetische kaart bevatte 531 merkers in 28 linkage groepen. De totale genetische afstand van de linkage groepen was 2.324 centimorgans (cM), met de grootste linkage groep (81 merkers) van 326 cM. Het gemiddelde merker interval voor alle merkers in de 28 linkage groepen was 4,6 cM. Het vergelijken van de genetische kaart van kalkoenen en kippen bevestigde twee inter-, en 57 intra-chromosomale herschikkingen tussen deze twee pluimveesoorten die tot de conclusie leiden dat de structuur van het genoom van de kalkoen en kip sterk geconserveerd is met een relatief laag aantal inter- en intra-chromosomale herschikkingen.

In **hoofdstuk 4** hebben we de genetische kaart van kalkoen, die werd geconstrueerd in hoofdstuk 3 van dit proefschrift, gebruikt samen met de individuele fenotypes om quantitative trait loci (QTL) in kaart te brengen in dezelfde populatie voor verschillende groeicurve kenmerken, lichaamsgewicht, borstvlees opbrengst en vlees kwaliteit kenmerken. Resultaten toonden QTL op 21 van de 27 kalkoen chromosomen van de genetische kaart. Voor alle kenmerken te samen werden 45 QTL gevonden in 29 verschillende regio's op de 21 chromosomen. Van de 45 QTL, waren er twaalf significant (chromosome wide P <0,01), terwijl de resterende 33 suggestief waren (chromosoom-wijde P <0,05) voor verschillende groei kenmerken, groeicurve, borstvlees opbrengst en vleeskwaliteit kenmerken. Deze studie suggereert pleiotropie of sterke linkage tussen QTL voor een aantal van de chromosomale regio's. Een vergelijking tussen de locatie van QTL ontdekt in kalkoen en de gerapporteerde kip QTL (QTLdb), op de syntenic kip chromosomen voor dezelfde of vergelijkbare kenmerken, liet tekenen van functionele conservatie van deze pluimveesoorten zien.

De volgende stap na het verkennen van de genetica van kalkoen productie (schatting van genetische parameters, constructie van de genetische kaart en de

172

QTL kaart) met behulp van de referentiepopulatie was om de variatie in het genoom van de kalkoen te onderzoeken (SNP ontdekking, populatie diversiteit en tekenen van selectie). **Hoofdstukken 5 en 6** beschrijven het gebruik van next generation sequencing technieken om variatie en tekenen van selectie in het genoom van de kalkoen te onderzoeken. Next generation sequencing werd uitgevoerd op 32 individuen uit elf verschillende kalkoen populaties (zeven commerciële, drie heritage en een Zuid-Mexicaanse wilde populatie). **Hoofdstuk 5** beschrijft hoe het hele genoom werd gescreend om SNPs te ontdekken in het kalkoengenoom, wat resulteerde in 5,49 miljoen mogelijke SNPs in vergelijking met het referentiegenoom. Vervolgens werd de genetische diversiteit tussen de verschillende populaties onderzocht met behulp van de ontdekte SNPs. In totaal werden 75.254 SNPs ontdekt in exon regio's, waarvan 23.795 niet-synonieme, 52.506 synonieme, 377 stop-gain en 8 stop-loss SNPs. De gemiddelde frequentie van heterozygote nucleotide posities in individuele kalkoenen was 1,07 $Kb^{-1}$. Dit niveau van heterozygositeit in het kalkoengenoom is laag in vergelijking met andere diersoorten, zoals kip en varken, en wijst op veel minder diversiteit in het kalkoengenoom. Lage heterozygositeit tussen commerciële lijnen, evenals de aanwezigheid van allelen/haplotypes die enkel in de wilde Zuid-Mexicaanse populatie zijn gevonden en niet in gedomesticeerde populaties, onderstreept dat er geselecteerd is voor specifieke haplotypes of dat deze verloren zijn gegaan in de moderne gedomesticeerde kalkoen. In een genetische diversiteit analyse vertakte alle commerciële lijnen vanaf één knooppunt ten opzichte van de erfgoed rassen en de wilde kalkoen populatie, wat aangeeft dat de commerciële lijnen van een gemeenschappelijke voorouder afstammen.

In **hoofdstuk 6**, werd de next generation sequencing data van tien kalkoen populaties (29 individuals) gebruikt om selective sweep regio's te detecteren. Over de kalkoen populaties heen zijn 54 regio's waargenomen met een significant (P <0,05) bewijs voor een selective sweep. Deze sweeps waren verdeeld over 14 verschillende chromosomen. Van deze 54 significante selective sweep regio's zijn er 31 populatie specifiek en overlappen er 23 met een selective sweep gebied in één of meer andere populaties. De 23 overlappende selective sweep regio's waren verdeeld over 13 verschillende kalkoen chromosomen. Van de 31 populatie specifieke selective sweep regio's, werden 26 gevonden in de commerciële populaties. De waargenomen selective sweep gebieden zijn groot. Het relatief hoge aantal selective sweep regio's in de commerciele kalkoen populaties in vergelijking met kalkoen erfgoed rassen, en de verrijking van kalkoen selective sweep regio's met genen die van belang zijn voor groei, geeft aan dat de selective sweep regio's

in kalkoen waarschijnlijk het gevolg zijn van intensieve selectie voor de groei, waardoor specifieke haplotypes worden gefixeerd.

Tenslotte, bespreek ik in **hoofdstuk 7** de belangrijkste bevindingen van dit proefschrift met betrekking tot hun gevolgen voor fokkerij. In dit hoofdstuk worden de rol van genen en het genoom bij het beheersen van complexe economisch belangrijke kenmerken besproken. In dit hoofdstuk bediscussieer ik ook manieren/technieken om varianten in het genoom te identificeren die de prestaties van individuen voor specifieke kenmerken beïnvloeden. Het effect van domesticatie en selectie van kalkoenen op de diversiteit wordt beschreven. Verder beschrijf ik de huidige selectie procedures die worden toegepast in de kalkoen industrie, uitdagingen voor de fokkers bij het toepassen van genomic selection en hoe het gebruik van genomische informatie effectief kan zijn in de kalkoen industrie.

# Acknowledgements

**Acknowledgements**

# Curriculum Vitae

## About the author

Muhammad Luqman Aslam was born in Faisalabad district of Punjab province in Pakistan on 1st July 1982, and he grew up in Faisalabad city. He obtained his Higher Secondary School Certificate (FSc. Pre-Medical) from the board of intermediate and secondary education (BISE) Faisalabad in 2001. In that same year, he started his Bachelor [B.Sc (Hons)] study in Animal Husbandry at the University of Agriculture Faisalabad and completed this in the year of 2005. Immediately after the completion of B.Sc (Hons), he started his Master [M.Sc (Hons)] in Animal Breeding and Genetics with focus on molecular genetics at the same university. He completed his M.Sc (Hons) in July 2007, with thesis entitled "Genetic Diversity among Native Aseel Chicken Population. He used microsatellite genetic markers for this genetic diversity analysis. During his M.Sc (Hons), he also worked for one year from May, 2006 to May 2007 as a research assistant in ALP-PARC research project entitled "Genetic Characterization of Native Cattle and Buffalo Breeds of Pakistan". In the duration of M.Sc (Hons), he dealt with intensive lab work, in-depth courses and a thesis based on research conducted on native Aseel chicken population. In October 2007, he started his PhD programme at the Animal Breeding and Genomics Centre, Wageningen University, The Netherlands, which resulted this thesis.

## List of publications

**Peer-reviewed publications**

**Aslam M. L.,** J.W.M. Bastiaansen, M. G. Elferink, H-J Megens, R.P.M.A. Crooijmans, L. Blomberg, R.C. Fleischer, C.P. Van Tassell, T. S. Sonstegard, S. G. Schroeder, M.A.M. Groenen, J.A. Long. 2012. Whole genome SNP discovery and analysis of genetic diversity in turkey (*Meleagris gallopavo*). BMC Genomics (In Press).

**Aslam, M. L.,** J.W.M. Bastiaansen, R.P.M.A. Crooijmans, B. J. Ducro, A. Vereijken, M.A.M. Groenen. 2011. Genetic variances, heritabilities and maternal effects on body weight, breast meat yield, meat quality and the shape of the growth curve in turkey birds. BMC Genetics. **12**: 14.

**Aslam, M. L.,** J.W.M. Bastiaansen, R.P.M.A. Crooijmans, A. Vereijken, M.A.M. Groenen. 2011. Whole genome QTL mapping for growth, meat quality and breast meat yield traits in turkey. BMC Genetics. **12**:61

**Aslam, M. L.,** J.W.M. Bastiaansen, R.P.M.A. Crooijmans, A. Vereijken, H-J. Megens, M.A.M. Groenen. 2010. A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the Turkey and Chicken genomes. BMC Genomics. **11**:647.

Dalloul, R. A., J. A. Long, A.V. Zimin, **Luqman Aslam,** et al., 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris Gallopavo*): genome assembly and analysis. Plos Biology. 8(9): e1000475.

**Publications in review or submitted**

**Aslam M. L.,** J.W.M. Bastiaansen, M. G. Elferink, H-J Megens, R.P.M.A. Crooijmans, L. Blomberg, C.P. Van Tassell, T. S. Sonstegard, S. G. Schroeder, J.A. Long, M.A.M. Groenen. 2012. Genome-wide signatures of selection revealed through massive parallel sequencing of DNA across ten turkey populations (*Meleagris gallopavo*). BMC Genetics (Submitted).

*Conference proceedings*

**Aslam, M. L.,** J.W.M. Bastiaansen, M. Elferink, H-J Megens, R.P.M.A Crooijmans, L. Blomberg, R. C. Fleischer, C. P. Van Tassell, T. S. Sonstegard, S. G. Schroeder, M.A.M. Groenen, J. A. Long. Selective sweep analysis in turkey using whole genome sequence data. *Book of abstracts* of the International Plant and Animal Genome Conference (PAG XX). San Diego, CA, USA (2012).

**Aslam**, M. L., J.W.M. Bastiaansen, M. Elferink, H-J. Megens, R.P.M.A. Crooijmans, L. Blomberg, R. C. Fleischer, C.P. Van Tassell, T. S. Sonstegard, S. G. Schroeder, M.A.M.

Groenen, J. A. Long. SNP Discovery and Genetic Diversity of Turkey. *Book of abstracts* Wageningen Institute of Animal Sciences (WIAS), Science Day. The Netherlands (2012).

**Aslam, M. L.**, J.W.M. Bastiaansen, R.P.M.A. Crooijmans, A. Vereijken, M.A.M. Groenen. QTL mapping for the logistic growth curve traits in turkeys. *Book of abstracts* of the International Plant and Animal Genome Conference (PAG XIX). San Diego, CA, USA (2011).

**Aslam, M. L.**, J.W.M. Bastiaansen, R.P.M.A. Crooijmans, A. Vereijken, M.A.M. Groenen. Whole genome QTL mapping for growth, breast yield and meat quality traits in turkey. *Book of abstracts* of the International Plant and Animal Genome Conference (PAG XIX). San Diego, CA, USA (2011).

**Aslam, M. L.,** J.W.M. Bastiaansen, R.P.M.A. Crooijmans, B. J. Ducro, A. Vereijken, and M.A.M. Groenen. 2010. Estimation of genetic parameters for the logistic growth curve in turkeys. Proceedings of $9^{th}$ World Congress on Genetics and Applied to Livestock Production Leipzig Germany (WCGALP, 2010).

**Aslam, M. L.,** J.W.M. Bastiaansen, R.P.M.A. Crooijmans, A. Vereijken, H-J. Megens, M.A.M. Groenen. 2010. Linkage map of the turkey genome reveals genomic rearrangements between turkey and chicken. *Book of abstracts* of the annual ZonMw, Genetica Retraite, Kerkrade, Maastricht, The Netherlands (2010).

**Training and Supervision Plan**

**The Basic Package (3 ECTS)**

| | |
|---|---|
| WIAS Introduction Course | 2008 |
| Course on Philosophy of Science and/or Ethics | 2008 |

**Scientific Exposure** (9 ECTS) *conferences, seminars and presentations*

| | |
|---|---|
| Animal Disease Genomics: opportunities and applications, Edinburgh | 2008 |
| EADGENE Ethical Matrix Workshop, Wageningen, The Netherlands | 2008 |
| QTLMAS Workshop, Wageningen, The Netherlands | 2008 |
| WIAS Science Conference, Wageningen, The Netherlands | 2008 |
| WIAS Science Conference, Wageningen, The Netherlands | 2009 |
| Genetica Retraite Maastricht, The Netherlands | 2010 |
| World Congress on Genetics Applied to Livestock Production | 2010 |
| WIAS Science Conference, Wageningen, The Netherlands | 2010 |
| Plant and Animal Genome XIX | 2011 |
| WIAS Science Conference Wageningen, The Netherlands | 2012 |

**In-Depth Studies** (18 ECTS)

| | |
|---|---|
| Mathematical Modelling in Biology | 2008 |
| Introduction to  R  for Statistical Analysis | 2008 |
| Genomics | 2008 |
| Gene Technology | 2008 |
| Getting Started with ASReml | 2009 |
| Design of Experiments | 2010 |
| Statistics for Life Sciences | 2010 |

**Professional Skills Support Courses** (5 ECTS)

| | |
|---|---|
| Dutch Language Course Level 1 | 2008 |
| Dutch language course level 11 | 2009 |
| Dutch language course level 111 | 2010 |

**Didactic Skills Training** (2 ECTS)

| | |
|---|---|
| RMC Course Reviewer | 2010 |
| Supervising MSc Student | 2011 |

| | |
|---|---|
| **Total Credits (ECTS)** | **36.8** |

**Colophon**

**Colophon**