Selection signature detection in a diverse set of chicken breeds

Dissertation

to obtain the Ph. D. degree

in the International Ph. D. Program for Agricultural Sciences in Goettingen (IPAG)

at the Faculty of Agricultural Sciences,

Georg-August-University Göttingen, Germany

presented by

Mahmood Gholami

born in Tehran

Göttingen, September 2014

D7

1st Referee: Prof. Dr. Henner Simianer
Animal Breeding and Genetics Group
Department of Animal Sciences
Georg-Agust-University Göttingen

2nd Referee: Dr. Steffen Weigend
Institute of Farm Animal Genetics (ING)
Friedrich-Loeffler-Institut (FLI)

Date of dissertation: 20.11.2014

TABLE OF CONTENTS

SUMMARY

Over the last decade, interest in detection of genes or genomic regions that are targeted by selection has been growing. Identifying signatures of selection can provide valuable insights about the genes or genomic regions that are or have been under selection pressure, which in turn leads to a better understanding of genotype-phenotype relationships. The main focus of this thesis is the detection of selection signatures in various breeds of chicken. Common strategy for the detection of selection signatures is to compare samples from several populations and search for genomic regions with outstanding genetic differentiation. This strategy uses inter-populations statistics. In this dissertation in each chapter (chapter 2, 3 and 4) one or two inter-populations statistics for selection signature detection are investigated.

Two sets of data set were used in this thesis: The first set comprised a total of 96 individuals of three commercial layer breeds (White Leghorn, White Rock and Rhode Island Red), 12 non-commercial fancy breeds and two subspecies of Junglefowls (*G. g. gallus* and *G. g. spadiceus*) were genotyped with three different 600K SNP-chips. The second set comprised pool sequences (10 individuals per pool) from 43 different chicken breeds. Including 3 commercial breeds (White Leghorn, White Rock and Rhode Island Red), 37 non-commercial breeds and three subspecies of Junglefowls (*G. g. gallus*, *G. g. spadiceus* and *G. varius*).

In our first approach, as described in the ***2nd chapter***, Wright's fixation index, F_{ST}, was used as an index of genetic differentiation between populations for detection of selection signatures on the first data set. This chapter focuses on detection of selection signatures between different chicken groups based on SNP-wise F_{ST} calculation. After removing overlapping SNPs between the three 600K SNP arrays a total of 1,139,073 SNPs remained. After filtering for minor allele frequencies lower than 5% and removing SNPs on unknown locations, a total of ~1 million SNPs were available for F_{ST} calculation. The average of F_{ST} values were calculated for overlapping windows. Average F_{ST} values between overlapping windows were then compared to detect for selection signatures. Two sets of comparisons were made in this study in order to detect selection signatures. First, we performed a comparison between commercial egg layers and non-commercial breeds and second within commercial egg-layer (white egg layers and brown egg layers). Comparing non-commercial and commercial breeds resulted in the detection of 630 selection signatures, while 656 selection signatures were detected in the comparison between commercial egg-layer breeds. Annotation of selection signature regions revealed various genes corresponding to productions traits, for which layer

breeds had been selected. *NCOA1*, *SREBF2* and *RALGAPA1* were among the detected genes, which are associated with reproductive traits, broodiness and egg production. Several of the detected genes were associated with growth and carcass traits, including *POMC*, *PRKAB2*, *SPP1*, *IGF2*, *CAPN1*, *TGFb2* and *IGFBP2*. These genes are good candidates for further studies. Our approach in chapter 2 demonstrates that including different populations with specific breeding histories can provide a unique opportunity for a better understanding of farm animal selection.

In the study described in the ***3rd chapter***, our aim was to use haplotype frequencies and considering the hierarchical poupolation structure in order to detect selection signatures. We used a subset of the first data set with a total of 74 individuals of three commercial layer breeds (White Leghorn, White Rock and Rhode Island Red) and two subspecies of Junglefowls (*G. g. gallus* and *G. g. spadiceus*). To facilitate this, we used the statistical methods FLK and hapFLK. FLK calculates variation of the inbreeding coefficient by using a population's kinship matrix to incorporate hierarchical structure. A similar statistic is used in hapFLK but haplotype frequencies are used instead of allele frequencies. FLK and hapFLK were calculated in all layer breeds, using subspecies of Junglefowls individuals for the estimation of the ancestral genetic distance. FLK and hapFLK were applied to three groups; all layers, white layers and brown layers. A total of 107 and 41 regions were detected as selective signatures in the FLK and hapFLK studies, respectively. Annotation of selection signature regions revealed various genes and QTL corresponding to productions traits, for which layer breeds were selected. A number of the detected genes were associated with growth and carcass traits, including *IGF-1R*, *AGRP* and *STAT5B*. We also annotated an interesting gene associated with dark brown plumage mutational phenotype in chickens (*SOX10*). Our new analysis in chapter 3 provided a great comparison between F_{ST}, FLK and hapFLK. Large overlap exists between the regions that have been determined as regions under selection in the F_{ST} study and in the current analysis using FLK and hapFLK. QTL associated to meat production as well as both *IGF-1R* and *STAT5B* were located in regions that were similar between brown layers. These results showed a large degree of agreement with the F_{ST} results discussed in chapter 2. We demonstrated that using haplotype frequencies and considering a hierarchical structure can improve the power of detection in our data set.

The approach discussed in the ***4th chapter*** of this dissertation uses SNPs extracted from the pool sequence data (the second data set) to detect selection signatures. Over 30 million SNPs in 43 pools consisted of 3 commercial breeds (White Leghorn, White Rock and Rhode Island Red), 37 non-commercial breeds and three subspecies of Junglefowls (*G. g. gallus*, *G. g.*

spadiceus and *G. varius*) were used in this study. After filtering for mapping quality and sequencing depth, 22 million SNPs remained. The breeds were studied for selection signature in three contrasts (i.e. skin color, egg shell color, and toe number). F_{ST} was calculated between the two groups, whereas heterozygosity (H_E) was obtained for each group. Both measures (F_{ST} and H_E) were subsequently summarized in 40 kb windows with an overlap of 50% within each contrast. Comparisons of summarized F_{ST} and H_E between overlapping windows was employed for selection signature detection, this was done to improve the power and reliability of detection. A total of eight regions (in all contrast) were detected as selective signatures using both F_{ST} and H_E methods. Annotation of selection signature regions revealed one gene (*BCO2*) and three QTL corresponding to skin color and egg shell color, respectively. In this study we demonstrated that the use of sequence data with a larger number of populations and combination of methods with different statistic background (i.e. F_{ST} and H_E) can improve the power of detection.

In conclusion, the results of the studies discussed in this dissertation showed that the identification of regions that were potentially under selection can be carried out by including various populations and utilizing of high-resolution genome scans (using dense marker or pool sequence data). Our study provides a great comparison between different inter-populations methods for selection signature detection (F_{ST}, FLK and hapFLK) and the use of different resolution of genome scans (pool sequence and high density chip). It is demonstrated that use of inter-populations (F_{ST}) method with combination of an intra-populations statistic (heterozygosity) can improve the power of detection. Several putative selection signature regions with genes corresponding to the productions traits that chicken breeds were selected for were identified in this study. These regions are good candidates for further studies for both commercial purposes and biodiversity studies. This study gives a better understanding of farm animal selection, particularly in regard to chicken.

1st Chapter

General introduction

General introduction

The main focus of this thesis is the detection of selection signatures in various breeds of chicken. In this introductory chapter different aspect of chicken (i.e. chicken genome, chicken domestication and chicken breeding) as a common domestic animal is discussed. In this thesis, a comparison between different methods for selection signature detection and the use of different resolution of genome scans (pool sequence and high density chip) is conducted. Therefore, a short overview of different methods in selection signature detection and availability of data for chicken are described in this chapter as well. Finally, relevant studies of selection signature detection in chicken are reviewed.

Chicken as farm animals

Chicken (*Gallus gallus domesticus*) is a domesticated fowl, that is one of the most common and widespread domestic animals (Perrins, 2003); there are more chickens in the world than any other species of bird. More than 50 billion chickens are raised annually as a source of food, for both their meat and eggs (“About chickens | Compassion in World Farming,” n.d.). Chickens raised for eggs are usually called layers, while the ones raised for meat are often called broilers. Chicken meat and eggs provide a leading source of high quality protein at a time when worldwide demand for this source of nutrition is growing rapidly (Rosegrant and Cai, 2001). Beyond the importance of a safe and nutritious food supply to human life, the enormous world-wide interest in raising poultry for food provides a collateral source of scientific data that expands our understanding of biology in general. The commercial populations enable large scale breeding studies on the chicken with unprecedented genetic resolution.

Chicken genome

Chicken has a compact genome compared with mammals, averaging about 1.2 Gb in size, with 39 diploid chromosomes (2n = 78). Chicken chromosomes are classified as five pairs of macrochromosomes, five pairs of intermediate chromosomes, twenty-eight pairs of microchromosomes and two sexual chromosomes (Groenen et al., 2000). Microchromosomes represent approximately one third of the total genome size, and have been found to have a much higher gene density than macrochromosomes. It is estimated that the majority of genes

in chicken genome are located on microchromosomes (Burt, 2002a). The most genetic diversity of any chromosome in chicken was also found to be on chromosome 16 (microchromosomes) (Wong et al., 2004), which is due to existence of major histocompatibility complex (MHC) (Hála et al., 1981) on this chromosome.

Sex chromosome of chicken - and birds in general- , named chromosome Z and W, are different from mammalian sex chromosomes. Male chickens are homogametic (ZZ), while females are heterogametic (ZW) (Nam and Ellegren, 2008). In chicken the ovum (egg cell) determines the sex of the offspring and the Z chromosome is larger and has more genes in contrast to W chromosome, similar to the X chromosome in the XY system (Bellott et al., 2010).

Chicken domestication

The domestic chicken is descended primarily from the Red Junglefowl (*Gallus gallus*) and is scientifically classified as the same species (Wong et al., 2004). Although Darwin suggested that the domestic chicken is descended from a single original species in Southeast Asia (Darwin, 1868), new studies suggested that the origin of domestic chickens lies in multiple origins in India and South-East Asia nearly 10,000 years ago (Tixier-Boichard et al., 2011). The debate of single and multiple origins of domestic chicken has been going on for decades. Eriksson *et al.* (2008) reveal that at least the gene for yellow skin was introuced into domestic birds through hybridization with the Grey Junglefowl (*Gallus sonneratii*). Kanginakudru *et al.* (2008) found evidence for domestication of chicken from two Red Junglefowl subspecies (*G. g. spadiceus* and *G. g.* gallus), as well as from Indian red jungle fowl (*G. g. murghi*) in the Indus valley. Liu *et al.* (2006) suggested different origins from different regions, such as Yunnan, South and Southwest China and the Indian subcontinent. Nishibori *et al.* (2005) indicated that inter-species hybridizations have occurred between Junglefowl and Ceylon Junglefowl (*Gallus lafayetii*).

Chicken breeding

Selective breeding of chicken has been documented as early as Roman times (Crawford, 1990). However, in contrast to current worldwide consumption of chicken meat and eggs as the major protein source (Al-Nasser et al., 2007), chicken may have been domesticated for cultural purposes such as religion, decoration, and cock fighting, rather than for food production (Rose, 1996). Strong selection of production traits started in the 20th century when commercial breeds were selected for either egg-laying or meat production (Burt, 2005).

During the past 85 years, modern selective breeding has made spectacular progress in both egg and meat production traits (Burt, 2002b). During this period, egg production (number of eggs per hen per year) has increased three-fold and growth rate (days to 1.5 kg live weight) four fold (Burt, 2002b). Currently, chicken plays an important rule as one of the major protein source for human. Currently world egg production has increased to 60 millions of tons and broiler meat to 90 millions of tons ("FAO Statistical Yearbook 2013," n.d.).

Data availability

A large amount of genomic information is already publicly available for chicken, including the first draft of reference genome (International Chicken Genome Sequencing Consortium, 2004), more than 3.5 million genetic variants in public databases (Sherry et al., 2001), and more than 3,000 QTLs in the Chicken QTLdb (Hu et al., 2013). The current reference genome of chicken (Gallus_gallus-4.0 released in 2011) contains 29 of 39 chromosomes in chicken with two linkage group and both sex chromosomes. Development of high density 600K SNP genotyping array for chicken (Kranis et al., 2013) and low cost of whole genome re-sequencing (Bentley, 2006) has facilitated high throughput investigation of many individuals for research and commercial application such as in genomic selection, genome-wide association studies, selection signature analyses, fine mapping of QTL, and detection of copy number variants.

Selection signature

''Selection signatures'' are defined as regions of the genome that contain a beneficial mutation and therefore are or have been under natural or artificialselection, leaving special patterns of DNA behind (Qanbari and Simianer, 2014). A local reduction of genetic variation up- and downstream of the beneficial mutation is caused by the rapid fixation of a beneficial mutation (Figure 1), leaving special patterns of DNA behind (Smith and Haigh, 1974). Selective sweep is the process by which a new beneficial mutation eliminates or reduces variation in linked neutral sites as it increases in frequency in the population (Braverman et al., 1995). The classic model of positive selection states that selection acts upon a newly arisen advantageous mutation, so that there is only one founding haplotype at the time of selection. Alternatively, selection could act on preexisting genetic variation that was previously either neutral or deleterious, but has become adaptive due to changes in the environment or genetic background (Akey, 2009). Selection from standing variation has been referred to as a ''soft sweep'' (Hermisson and Pennings, 2005), to distinguish it from the

classic model, or “hard sweep”. The selection signatures can be used to screen a genome for genes involved in recent adaptation.

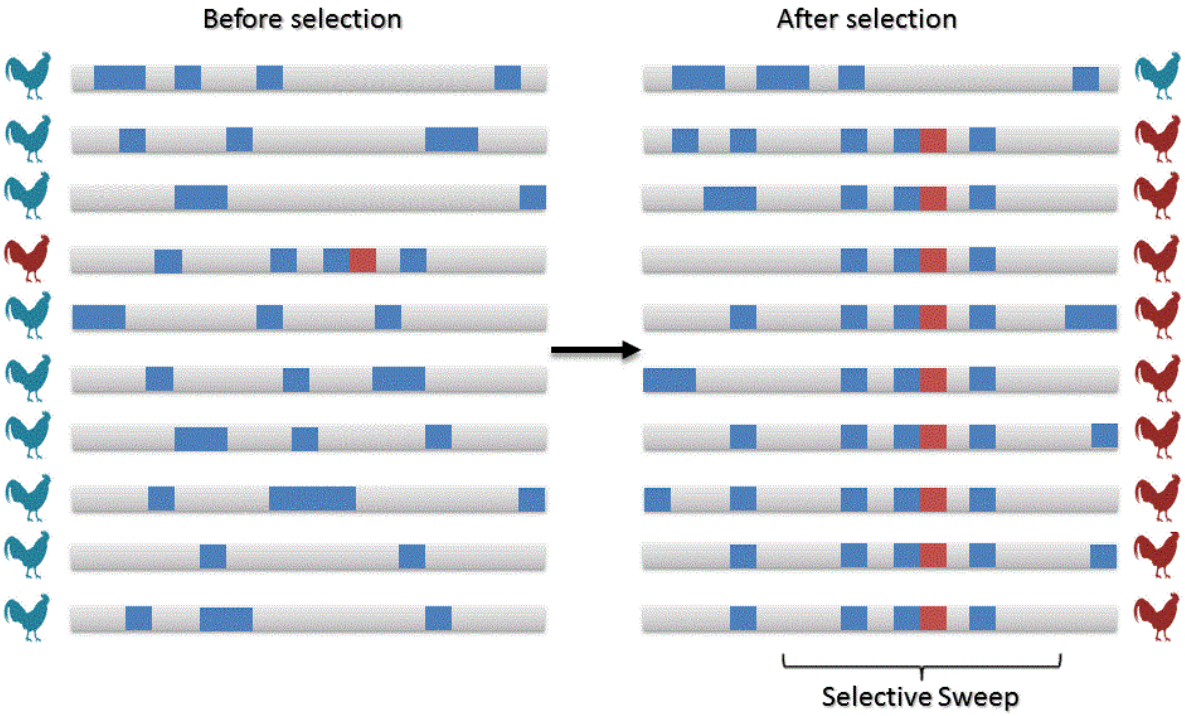

Figure 1. Reduction of genetic variation up- and downstream of the beneficial mutation (in red) is caused by the rapid fixation of a beneficial mutation due to selection.

Methods of selection signature detection

Based on the hitchhiking theory positive selection can leave a set of informative signatures (i.e., reduced local variability, deviated spectrum of allele frequencies and a specific linkage disequilibrium pattern). Based on these signatures, a variety of statistical approaches are available for selection signature detection from SNP data (SNP-chip or sequence data). Qanbari *et al.* (2014) classified these methods in two main groups: intra-population statistics and inter-populations statistics.

Intra-population statistics searches for informative signatures by comparison of genomic data within populations. Intra-population statistics are focused on three neutrality theory:

i. Site frequency spectrum (SFS) is a class of tests which summarizes the allele frequency distribution of polymorphisms in a region of interest. A widely used statistic established in this class is Tajima's D (Tajima, 1989). A more recent statistic in this class is the maximum of composite likelihood ratio (CLR) (Nielsen et al., 2005).

ii. Linkage disequilibrium (LD) refers to the nonrandom association of alleles between two or more loci. An ongoing or incomplete selection signature has a high-frequency haplotype with extended LD, which is mainly because recombination dose not (or rarely) occur during the rapid increase in frequency of a haplotype carrying a beneficial mutation. Popular LD based tests include relative extended haplotype homozygosity (rEHH) (Sabeti et al., 2002), integrated haplotype score (iHS) (Voight et al., 2006) and linkage disequilibrium decay test (LDD) (Wang et al., 2006).

iii. Reduced local variability is a class of methods that identify genomic regions with a systematically reduced variation (e.g., nucleotide diversity or heterozygosity) relative to the average across the genome. Some tests in this class are runs of homozygosity (ROH) (McQuillan et al., 2008) and pooled heterozygosity (H_P) (Rubin et al., 2010).

Inter-populations statistics compare genomic data between two or more populations to identify regions with informative signatures. Statistics in these methods focus on differentiation between populations. According to the theory that most populations exhibit some degree of population structure, compression of genomic data between populations can reveals regions that have been under selection in different populations. Statistics in this group can be classified in two groups:

i. Single site differentiation is the simplest and most popular group. The statistic used to detect local increases in population under selection is F_{ST} (Wright, 1949). A more novel statistic based on single site differentiation is FLK (Bonhomme et al., 2010).

ii. Haplotype based differentiation analyses; ascertainment bias of SNP has less effect using haplotype clusters rather than SNPs. Methods in this class use haplotype information in multiple population comparisons. One of the popular methods in this class is cross population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007). Another example is hapFLK, a haplotype based extension of the FLK statistic (Fariello et al., 2013).

In the following sections few of the inter-populations statistics and heterozygosity which were used in this study, are explained in more detail:

F_{ST}

Wright's fixation index, F_{ST}, is a useful index of genetic differentiation between populations (Wright, 1949). If $\bar{p}$ is the average frequency of an allele in the total population, σ_S^2 is the variance in the frequency of alleles in different subpopulations, and σ_T^2 is the variance of allele frequencies in the total population, F_{ST} is defined as:

$$F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\bar{p}(1-\bar{p})} \quad (1)$$

Other estimators of F_{ST} have been proposed as well, including a modern analogue for multi-allele loci known as Weir & Cockerham's F_{ST} Estimator (Weir and Cockerham, 1984) as well as a F_{ST} estimator with Bayesian model (Gianola et al., 2010).

F_{ST} reflects the degree of differentiation between populations at any given locus, ranging from 0 (no differentiation) to 1 (fixed difference between populations). Negative or balancing selection tends to decrease F_{ST}, whereas local positive selection tends to increase F_{ST} (Barreiro et al., 2008). Genes responsible for phenotypic differences between populations are expected to show large allele frequency differences (Myles et al., 2008). F_{ST} has an advantage over multi-locus testing such as SFS or LD based methods in that it is SNP-specific and can theoretically reveal the actual genetic variants under selection. However, since selective sweeps causes a whole series of SNPs to display an elevated F_{ST} profile, it is more efficient to look for a number of consecutive SNPs with average F_{ST} score (by use of genomic windows) rather than analyzing each SNP separately. F_{ST} is used in **chapters 2** and **4**.

FLK

FLK calculates variation of the inbreeding coefficient and incorporate hierarchical poupolation structure (Bonhomme et al., 2010). FLK is an extension of the original lewontin and krakauer (LK) statistic (Lewontin and Krakauer, 1973) that uses a phylogenetic estimation of the population's kinship (*F*) matrix, and, thus it deals with population effective size (*Ne*) variation and historical branching of populations. *F* matrix is a measure of the expected drift on each population and the expected covariance between them (for details read

Bonhomme et al., 2010). For FLK calculation, first Pzero (p_0) is estimated through the kinship matrix from the data, as follows:

$$\hat{p}_0 = \frac{1_n' F^{-1} p}{1_n' F^{-1} 1_n} \quad (2)$$

, where p is the allele frequencies for SNP, n is the number of populations and 1_n is an n-vector of 1's. Then, FLK is calculated as:

$$T_{FLK} = (p - \hat{p}_0 \mathbf{1}_n)' \widehat{Var(p)}^{-1} (p - \hat{p}_0 1_n) \quad (3)$$

, where $\widehat{Var(p)}^{-1}$ is the expected covariance matrix of vector *p*, which is estimated as:

$$\widehat{Var(p)} = F\hat{p}_0(1 - \hat{p}_0) \quad (4)$$

FLK is a parametric statistical test for detection of selection signatures in complex population trees. FLK is a quick and powerful tool for large data sets in the context of genomic scans. Bonhomme *et al.* (2010) showed that using FLK to detect selection signatures in comparison to other F_{ST}-like approaches (F_{ST} and LK statistic) greatly decreases the type one error (Bonhomme et al., 2010). FLK is used in **chapter 3**.

hapFLK

Haplotype diversity and LD patterns contain useful information for the detection of selection signatures (Sabeti et al., 2007) and therefore, usage of haplotype or LD based differentiation analyses has its own advantages. Most of the haplotype differentiation scans does not account for the possibility of hierarchical structure between populations. Fariello et al. (Fariello et al., 2013) proposed hapFLK statistic, which is a haplotype based extension of the FLK statistic (explained before) that accounts for both hierarchical population structure and haplotype information. The Scheet and Stephens model (Scheet and Stephens, 2006) summarizes local haplotype diversity in a sample through a reduction of dimension by clustering similar haplotypes together. These clusters can then be considered as alleles to compute the haplotype version of FLK statistic (for details read Fariello et al., 2013). Same kinship matrix (*F*) is used in hapFLK, but the statistic is computed from haplotype frequencies rather than SNP allele frequencies. hapFLK is the mean of $\tilde{T}'_{FLK}$ through all expectation maximization (EM) runs

(for details read Scheet and Stephens, 2006). $\tilde{T}'_{FLK}$ is calculated with a slight modification in equation (3):

$$\tilde{T}'_{FLK} = (p_l - p_0 1_n)'(IF)^{-1}(p - p_0 1_n) \quad (5)$$

, where p_o is calculated by equation (2), n is the number of populations, 1_n is an n-vector of 1's and p_l is the haplotype frequency at marker *l* and cluster *k*.

$$p_l = \left(p_{11}^l, \dots, p_{1n}^l, p_{21}^l, \dots, p_{2n}^l, \dots, p_{k1}^l, \dots, p_{kn}^l\right)' \quad (6)$$

Simulations showed that two features of hapFLK (i.e. the use of haplotype information and hierarchical structure of populations) significantly improves the detection power of selected loci, and that combining them in the hapFLK statistic provides even greater power (Fariello et al., 2013). Specifically, Fariello *et al.* (2013) demonstrated that the hapFLK statistic has more power in detecting soft sweeps, incomplete sweeps and sweeps occurring in several populations. hapFLK is used in **chapter 3**.

Heterozygosity

Heterozygosity is the presence of different alleles at one locus on homologous chromosomes. Based on the hitchhiking theory a reduction of local variability exists in a selective sweep (Kaplan et al., 1989); therefore a selective sweep should have a lower heterozygosity (higher homozygosity) than the average heterozygosity (homozygosity) of the genome. There are several methods which scan genome based on heterozygosity. Runs of homozygosity (ROH) searches for continuous parts of the genome without heterozygosity in the diploid state, and is used on a genome-wide scale to detect signals of past selection (Lencz et al., 2007). Pooled heterozygosity (H_P) uses allele counts (based on sequence reads) to calculate heterozygosity (Rubin et al., 2010). Based on Hardy–Weinberg principle (Hardy, 2003; Weinberg, 1908) expected Heterozygosity (H_E) can be calculated from allele frequencies as:

$$H_E = 2pq \quad (7)$$

, where p is the allele frequency of an allele at a diploid locus and q is the allele frequency of its alternative allele. Since:

$$p + q = 1 \quad (8)$$

Therefore heterozygosity (H_E) can be calculated as:

$$H_E = 2p(1-p) \quad (9)$$

To identify regions under selection Z transformed heterozygosity (zH_E) is calculated. The *Z* transformation produces comparability of several breed pools with differing average heterozygosity within the same frame work, because quintile-based thresholds can be applied more easily on normalized values (Rubin et al., 2010). zH_E is calculated as:

$$ZH_E = \frac{(H_E - \mu H_E)}{\sigma H_E} \quad (10)$$

, where μH_E is the mean of heterozygosity and σH_E is the standard deviation of heterozygosity. Compared to inter-population methods mentioned above, heterozygosity can be run in a single population or within a group of populations. Heterozygosity is used in **chapter 4**.

Relevant studies in chicken

The growing genomic resources, relatively rapid reproduction time, and existence of several inbred lines, together with strong agricultural interest make chicken an excellent model for studying the signatures of selection under artificial breeding conditions (Brown et al., 2003). In recent years several groups have studied the selection signature in chicken. Employing re-sequencing data in order to detect selection signatures, Rubin *et al.* (2010) run a genome scan by calculating Pooled Heterozygosity (H_P) in 40-Kb sliding windows between 9 different lines (four broiler lines, four layers lines and one Red Junglefowl). They used *BCO2* (yellow skin gene) as a proof of concept for their detection method. They detected 21 regions as being under selection, including a region on chromosome 5 at the locus encoding thyroid stimulating hormone receptor (*TSHR*). Using a 60k SNP assay, Johansson *et al.* (2010) scanned the genome by calculating observed homozygosity and probability of fixation in two chicken lines, where 50 generations of selection have resulted in a 9-fold difference in body weight. They detected 50 regions as fixed in the population due to selection. Elferink *et al.* (2012) genotyped 67 lines (including broilers, layers and three subspecies of Junglefowls) with a 58K SNP chip, and calculated H_P in 5 markers window size (~97 kb) in different breed groups for selection signature detection. They also used *BCO2* (yellow skin gene) to validate

their selective signature detection method. They detected 396 regions that show suggestive evidence of selection, 26 of these regions showed strong evidence of selection. They detected several genes and QTL with biological functions that can be linked to production traits in chicken. Qanbari *et al.* (2012) calculated H_P and used creeping windows with 40 Kb size for re-sequencing data of 15 brown-egg layers in order to detect selection signature. They detected 132 regions as selection signature, among these regions were regions including *TSHR*, insulin-like growth factor 1 (*IGF1*), and several other genes related to production traits in chicken. Two studies have used relative extended haplotype homozygosity (rEHH) for selection signature detection in chicken (Zhang et al., 2012; Li et al., 2012). Useing 60K SNP chip Zhang *et al.* (2012) studied the selection signature in two broiler chicken lines which were divergently selected for abdominal fat content (lean and fat). They detected 51 and 57 regions that were under selection for lean and fat lines, respectively, these regions included several genes and QTL associated with fatness. Li *et al.* (2012) as well used 60K SNP chip to perform genome-wide scan for selection signature detection in 385 White Leghorn hens. They presented a genome-wide map of LD extent and several genes and QTL associated with egg production, metabolism traits, and response to illumination in their study.

Objective of this thesis

Chicken meat and eggs are one of the major protein sources for human, furthermore chicken has been a popular model organism for at least 100 years (Stern, 2005). Therefore better understanding of the chicken genome as a commercial farm animal and as a model organism is crucial. As a farm animal, better understanding of the genome can lead to genes or genomic regions that are associated with beneficial traits. Identifying signatures of selection can provide valuable insights about the genes or genomic regions that are or have been under selection pressure. Once a gene or genomic region with an impact on phenotype is located, this information can be incorporated in breeding value estimation by marker-assisted breeding (Ribaut and Hoisington, 1998) or into a genomic prediction model which can exploit already existing knowledge of genetic architectures (Zhang et al., 2014). Additionally, a better understanding of the chicken genome can enhance the use of chicken as a model organism for biomedical research (Burt, 2007).

As discussed above several methods exist for selection signature detection, some of them with different resolution of genome scans are investigated in this thesis:

Chapter 2 studies F_{ST} as a method for selection signature detection in chicken. In this chapter, F_{ST} is calculated for one million SNPs in two comparisons in order to detect selection signatures in egg-layers. Three commercial egg-layer breeds, 12 non-commercial fancy breeds and two subspecies of Junglefowls are used for these comparisons.

Chapter 3 describes the effect of haplotype frequencies and consideration of hierarchical structure for selection signatures detection. For this purpose FLK and hapFLK are used to detect selection signatures in three commercial egg-layers. A comparison between F_{ST}, FLK and hapFLK is performed in this chapter.

Chapter 4 discusses the effect of high-resolution genome scans and large breed diversity on selection signature detection. Over 20 million SNPs in 43 pools from 43 different breeds were used in this chapter. F_{ST} and heterozygosity is calculated for three comparisons, and hence, a combination of inter- and intra-poupolation methods for selection signature detection is discussed in this chapter as well.

Chapter 5 includes a general discussion of the effect of different methods and different resolution of genome scans on selection signature detection. Critical issues of methods for selection signatures detections that are used in this thesis are discussed as well.

References

About chickens | Compassion in World Farming [WWW Document], n.d. URL http://www.ciwf.org.uk/farm-animals/chickens/ (accessed 7.2.14).

Akey, J.M., 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19, 711–722. doi:10.1101/gr.086652.108

Al-Nasser, A., Al-KHALAIFA, H., Al-Saffar, A., Khalil, F., Albahouh, M., Ragheb, G., Al-Haddad, A., Mashaly, M., 2007. Overview of chicken taxonomy and domestication. Worlds Poult. Sci. J. 63, 285. doi:10.1017/S004393390700147X

Barreiro, L.B., Laval, G., Quach, H., Patin, E., Quintana-Murci, L., 2008. Natural selection has driven population differentiation in modern humans. Nat. Genet. 40, 340–345. doi:10.1038/ng.78

Bellott, D.W., Skaletsky, H., Pyntikova, T., Mardis, E.R., Graves, T., Kremitzki, C., Brown, L.G., Rozen, S., Warren, W.C., Wilson, R.K., Page, D.C., 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 466, 612–616. doi:10.1038/nature09172

Bentley, D.R., 2006. Whole-genome re-sequencing. Curr. Opin. Genet. Dev., Genomes and evolution 16, 545–552. doi:10.1016/j.gde.2006.10.009

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., SanCristobal, M., 2010. Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. Genetics 186, 241–262. doi:10.1534/genetics.110.117275

Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H., Stephan, W., 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140, 783–796.

Brown, W.R.A., Hubbard, S.J., Tickle, C., Wilson, S.A., 2003. The chicken as a model for large-scale analysis of vertebrate gene function. Nat. Rev. Genet. 4, 87–98. doi:10.1038/nrg998

Burt, D.W., 2002a. Origin and evolution of avian microchromosomes. Cytogenet. Genome Res. 96, 97–112. doi:63018

Burt, D. W., 2002b. Applications of biotechnology in the poultry industry. Worlds Poult. Sci. J. 58, 5–13. doi:10.1079/WPS20020002

Burt, D.W., 2005. Chicken genome: current status and future opportunities. Genome Res. 15, 1692–1698. doi:10.1101/gr.4141805

Burt, D.W., 2007. Emergence of the Chicken as a Model Organism: Implications for Agriculture and Biology. Poult. Sci. 86, 1460–1471. doi:10.1093/ps/86.7.1460

Crawford, R.D., 1990. Poultry Breeding and Genetics. Elsevier, New York.

Darwin, C., 1868. The variation of animals and plants under domestication.: By Charles Darwin, ... In two volumes. With illustrations. John Murray, Albemarle Street, London.

Elferink, M.G., Megens, H.-J., Vereijken, A., Hu, X., Crooijmans, R.P.M.A., Groenen, M.A.M., 2012. Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds. PLoS ONE 7, e32720. doi:10.1371/journal.pone.0032720

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., Jensen, P., Andersson, L., 2008. Identification of the Yellow Skin Gene Reveals a Hybrid Origin of the Domestic Chicken. PLoS Genet 4, e1000010. doi:10.1371/journal.pgen.1000010

FAO Statistical Yearbook 2013 [WWW Document], n.d. Issuu. URL http://issuu.com/faooftheun/docs/syb2013issuu (accessed 7.2.14).

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M., Servin, B., 2013. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. Genetics 193, 929–941. doi:10.1534/genetics.112.147231

Gianola, D., Simianer, H., Qanbari, S., 2010. A two-step method for detecting selection signatures using genetic markers. Genet. Res. 92, 141–155. doi:10.1017/S0016672310000121

Groenen, M.A.M., Cheng, H.H., Bumstead, N., Benkel, B.F., Briles, W.E., Burke, T., Burt, D.W., Crittenden, L.B., Dodgson, J., Hillel, J., Lamont, S., de Leon, A.P., Soller, M., Takahashi, H., Vignal, A., 2000. A Consensus Linkage Map of the Chicken Genome. Genome Res. 10, 137–147.

Hála, K., Boyd, R., Wick, G., 1981. Chicken major histocompatibility complex and disease. Scand. J. Immunol. 14, 607–616.

Hardy, G.H., 2003. Mendelian proportions in a mixed population. 1908. Yale J. Biol. Med. 76, 79–80.

Hermisson, J., Pennings, P.S., 2005. Soft Sweeps Molecular Population Genetics of Adaptation From Standing Genetic Variation. Genetics 169, 2335–2352. doi:10.1534/genetics.104.036947

Holsinger, K.E., Weir, B.S., 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat. Rev. Genet. 10, 639–650. doi:10.1038/nrg2611

Hu, Z.-L., Park, C.A., Wu, X.-L., Reecy, J.M., 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 41, D871–D879. doi:10.1093/nar/gks1150

International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716. doi:10.1038/nature03154

Johansson, A.M., Pettersson, M.E., Siegel, P.B., Carlborg, Ö., 2010. Genome-Wide Effects of Long-Term Divergent Selection. PLoS Genet 6, e1001188. doi:10.1371/journal.pgen.1001188

Kanginakudru, S., Metta, M., Jakati, R.D., Nagaraju, J., 2008. Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. BMC Evol. Biol. 8, 174. doi:10.1186/1471-2148-8-174

Kaplan, N.L., Hudson, R.R., Langley, C.H., 1989. The “hitchhiking effect” revisited. Genetics 123, 887–899.

Kim, Y., Stephan, W., 2002. Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. Genetics 160, 765–777.

Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P., Hocking, P.M., Fife, M., Salmon, N., Fulton, J., Strom, T.M., Haberer, G., Weigend, S., Preisinger, R., Gholami, M., Qanbari, S., Simianer, H., Watson, K.A., Woolliams, J.A., Burt, D.W., 2013. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics 14, 59. doi:10.1186/1471-2164-14-59

Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T.V., Kane, J.M., Kucherlapati, R., Malhotra, A.K., 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc. Natl. Acad. Sci. 104, 19942–19947. doi:10.1073/pnas.0710021104

Lewontin, R.C., Krakauer, J., 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74, 175–195.

Li, D.F., Liu, W.B., Liu, J.F., Yi, G.Q., Lian, L., Qu, L.J., Li, J.Y., Xu, G.Y., Yang, N., 2012. Whole-genome scan for signatures of recent selection reveals loci associated with important traits in White Leghorn chickens. Poult. Sci. 91, 1804–1812. doi:10.3382/ps.2012-02275

Liu, Y.-P., Wu, G.-S., Yao, Y.-G., Miao, Y.-W., Luikart, G., Baig, M., Beja-Pereira, A., Ding, Z.-L., Palanichamy, M.G., Zhang, Y.-P., 2006. Multiple maternal origins of chickens: Out of the Asian jungles. Mol. Phylogenet. Evol. 38, 12–19. doi:10.1016/j.ympev.2005.09.014

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H., Wilson, J.F., 2008. Runs of Homozygosity in European Populations. Am. J. Hum. Genet. 83, 359–372. doi:10.1016/j.ajhg.2008.08.007

Myles, S., Tang, K., Somel, M., Green, R.E., Kelso, J., Stoneking, M., 2008. Identification and analysis of genomic regions with large between-population differentiation in humans. Ann. Hum. Genet. 72, 99–110. doi:10.1111/j.1469-1809.2007.00390.x

Nam, K., Ellegren, H., 2008. The Chicken (Gallus gallus) Z Chromosome Contains at Least Three Nonlinear Evolutionary Strata. Genetics 180, 1131–1136. doi:10.1534/genetics.108.090324

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566–1575. doi:10.1101/gr.4252305

Nishibori, M., Shimogiri, T., Hayashi, T., Yasue, H., 2005. Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius. Anim. Genet. 36, 367–375. doi:10.1111/j.1365-2052.2005.01318.x

Perrins, C., 2003. Firefly Encyclopedia of Birds. Firefly Books, Buffalo, N.Y.

Qanbari, S., Simianer, H., n.d. Mapping signatures of positive selection in the genome of livestock. Livest. Sci. 0. doi:10.1016/j.livsci.2014.05.003

Qanbari, S., Strom, T.M., Haberer, G., Weigend, S., Gheyas, A.A., Turner, F., Burt, D.W., Preisinger, R., Gianola, D., Simianer, H., 2012. A High Resolution Genome-Wide Scan for Significant Selective Sweeps: An Application to Pooled Sequence Data in Laying Chickens. PLoS ONE 7, e49525. doi:10.1371/journal.pone.0049525

Ribaut, J.-M., Hoisington, D., 1998. Marker-assisted selection: new tools and strategies. Trends Plant Sci. 3, 236–239. doi:10.1016/S1360-1385(98)01240-0

Rose, S.P., 1996. Principles of Poultry Science. CABI Publishing, New York.

Rosegrant, M.W., Cai, X., 2001. Water scarcity and food security: alternative futures for the 21st century. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. 43, 61–70.

Rubin, C.-J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., Andersson, L., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591. doi:10.1038/nature08832

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837. doi:10.1038/nature01140

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. doi:10.1038/nature06250

Scheet, P., Stephens, M., 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78, 629–644. doi:10.1086/502802

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311.

Smith, J.M., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23, 23–35. doi:10.1017/S0016672300014634

Stern, C.D., 2005. The chick; a great model system becomes even greater. Dev. Cell 8, 9–17. doi:10.1016/j.devcel.2004.11.018

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Tixier-Boichard, M., Bed'hom, B., Rognon, X., 2011. Chicken domestication: from archeology to genomics. C. R. Biol. 334, 197–204. doi:10.1016/j.crvi.2010.12.012

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A Map of Recent Positive Selection in the Human Genome. PLoS Biol 4, e72. doi:10.1371/journal.pbio.0040072

Wang, E.T., Kodama, G., Baldi, P., Moyzis, R.K., 2006. Global landscape of recent inferred Darwinian selection for Homo sapiens. Proc. Natl. Acad. Sci. U. S. A. 103, 135–140. doi:10.1073/pnas.0509691102

Weinberg, W., 1908. Uber den {N}achweis der {V}ererbung beim {M}enschen. Jahresh. Ver. F Vaterl Naturk Wurttemberg 64, 368–82.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38, 1358. doi:10.2307/2408641

Wong, G.K.-S., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Zhang, J., Li, R., Li, S., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruan, J., Zhou, Y., Li, Y., He, X., Zhang, Y., Wang, J., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G., Dong, L., Lan, F., Sun, Y., Zhang, Z., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Wang, J., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L.W., Yang, S.-P., Warren, W.C., Wilson, R.K., Brandström, M., Ellegren, H., Crooijmans, R.P.M.A., van der Poel, J.J., Bovenhuis, H., Groenen, M.A.M., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J.R., Rogers, S., Walker, B.A., van Hateren, A., Kaufman, J., Bumstead, N., Lamont, S.J., Zhou, H., Hocking, P.M., Morrice, D., de Koning, D.-J., Law, A., Bartley, N., Burt, D.W., Hunt, H., Cheng, H.H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M.T., Andersson, B., Webber, C., Ponting, C.P., Overton, I.M., Boardman, P.E., Tang, H., Hubbard, S.J., Wilson, S.A., Yu, J., Wang, J., Yang, H., International Chicken Polymorphism Map Consortium, 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432, 717–722. doi:10.1038/nature03156

Wright, S., 1949. The Genetical Structure of Populations. Ann. Eugen. 15, 323–354. doi:10.1111/j.1469-1809.1949.tb02451.x

Zhang, H., Wang, S.-Z., Wang, Z.-P., Da, Y., Wang, N., Hu, X.-X., Zhang, Y.-D., Wang, Y.-X., Leng, L., Tang, Z.-Q., Li, H., 2012. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC Genomics 13, 704. doi:10.1186/1471-2164-13-704

Zhang, Z., Ober, U., Erbe, M., Zhang, H., Gao, N., He, J., Li, J., Simianer, H., 2014. Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies. PLoS ONE 9, e93017. doi:10.1371/journal.pone.0093017

2nd Chapter

Population genomic analyses based on 1 million SNPs in commercial egg layers

Mahmood Gholami[1], Malena Erbe[1], Christian Gärke[1], Rudolf Preisinger[2], Annett Weigend[3] , Steffen Weigend[3]and Henner Simianer[1]

1 Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Göttingen, Göttingen, Germany

2 LOHMANN Tierzucht GMBH, Cuxhaven, Germany

3 Institute of Farm Animal Genetics (ING), Friedrich-Loeffler-Institut (FLI), Neustadt, Germany

Published in PLOS ONE:

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0094509

Abstract

Identifying signatures of selection can provide valuable insight about the genes or genomic regions that are or have been under selective pressure, which can lead to a better understanding of genotype-phenotype relationships. A common strategy for selection signature detection is to compare samples from several populations and search for genomic regions with outstanding genetic differentiation. Wright's fixation index, F_{ST}, is a useful index for evaluation of genetic differentiation between populations. The aim of this study was to detect selective signatures between different chicken groups based on SNP-wise F_{ST} calculation. A total of 96 individuals of three commercial layer breeds and 14 non-commercial fancy breeds were genotyped with three different 600K SNP-chips. After filtering a total of 1 million SNPs were available for F_{ST} calculation. Averages of F_{ST} values were calculated for overlapping windows. Comparisons of these were then conducted between commercial egg layers and non-commercial fancy breeds, as well as between white egg layers and brown egg layers. Comparing non-commercial and commercial breeds resulted in the detection of 630 selective signatures, while 656 selective signatures were detected in the comparison between the commercial egg-layer breeds. Annotation of selection signature regions revealed various genes corresponding to productions traits, for which layer breeds were selected. Among them were *NCOA1*, *SREBF2* and *RALGAPA1* associated with reproductive traits, broodiness and egg production. Furthermore, several of the detected genes were associated with growth and carcass traits, including *POMC*, *PRKAB2*, *SPP1*, *IGF2*, *CAPN1*, *TGFb2* and *IGFBP2*. Our approach demonstrates that including different populations with a specific breeding history can provide a unique opportunity for a better understanding of farm animal selection.

Introduction

Charles Darwin suggested that the domestic chicken is descended from a single original species, the Red Jungle fowl (*Gallus gallus*), and that this happened in Southeast Asia nearly 10,000 years ago[2]. On the contrary, new studies suggested that the origin of domestic chickens lies in multiple origins in South and Southeast Asia [2, 3]. Selective breeding of chicken has been documented as early as Roman times. However, in contrast to current worldwide consumption of chicken meat and eggs as the major protein source [5] chicken may have been domesticated for cultural purposes such as religion, decoration, and cock fighting rather than for food production [6]. Strong selection of production traits started in the

20th century when commercial breeds were selected for either egg-laying or meat production [7].

Strong selection has a direct effect on nucleotide diversity. Reduction or loss of nucleotide diversity at and near the selected locus caused by strong selection on desirable alleles is often referred to as genetic hitch-hiking or as a selective sweep [8]. Studying such signatures of selection can provide valuable insights about the genes or genomic regions that are or have been under selective pressure and hence can help in understanding important genotype-phenotype relationships. The discovery of a massive number of single nucleotide polymorphisms (SNPs) in the genomes of several species has enabled exploration of genome-wide signatures of selection via an assessment of variation in marker allele frequencies among populations [9]. A common strategy in this context is to compare samples from several populations, and look for genomic regions with outstanding genetic differentiation. Wright's fixation index, F_{ST}, is a useful index of genetic differentiation between populations [10] and reflects the degree of differentiation between populations at any given locus, ranging from 0 (no differentiation) to 1 (fixed difference between populations). Negative or balancing selection tends to decrease F_{ST}, whereas local positive selection tends to increase F_{ST} [11]. Genes responsible for phenotypic differences between populations are expected to show large allele frequency differences [12].

The growing genomic resources, the relatively rapid reproduction time and the existence of several inbred lines together with strong agricultural interest makes chicken an excellent model for studying the signatures of selection under artificial conditions [13]. A number of recent studies have investigated selection signatures in chicken either using sequence data or genotype data from low to medium density SNP chips. For example, Rubin *et al.* [14] studied the signatures of domestication and selective sweeps in various commercial broiler and layer lines using Next Generation Sequencing data from pooled DNA samples by searching for genomic regions with high degree of fixation of alleles. Johansson *et al.* [15] used a 60K SNP chip to study the genome wide effect of divergent selection between two chicken lines with a 9-fold difference in body weight. Elferink *et al.* [16] studied selective sweeps using the same method described by Rubin *et al.* [14] but carried out the study on a large number of chicken breeds (67 in total) using a 58K SNP chip.

In this study, 96 individuals from three commercial layer breeds and 14 non-commercial fancy breeds, including Red Jungle fowl (Cochin-Chinese) (*G. g. gallus*) and Red Jungle fowl (Burmese) (*G. g. spadiceus*), were genotyped with three different 600K SNP-chip from

Affymetrix (with substantial proportion of overlapping SNPs between the three chips). This data set was produced during the validation of pre-screening arrays of the newly developed Axiom® Genome-Wide Chicken Genotyping Array [17]. Wright's fixation index, F_{ST}, was used to study signatures of selection in the large dataset. The analysis of this large dataset provides an excellent basis for detecting selection signatures in the genomes of the chicken breeds under study and is unprecedented regarding the combination of number of genotyped individuals and marker density applied. This in turn can provide important information on the genomic regions which have been under selection and associated with specific layer traits.

Material and methods

Animals, data collection and filtering

Two sets of samples, commercial egg layers and non-commercial fancy breeds (coded respectively LY and OG), were used for this study. The commercial individuals from Lohmann Tierzucht GmbH originated from three breeds: One commercial white egg layer breed based on White Leghorn (WL) with three separate lines and two brown egg layer breeds based on White Rock (WR) and Rhode Island Red (RIR), respectively, with two separate lines per breed. In each of these lines (seven in total) ten individuals were sampled and genotyped. The non-commercial fancy breeds consist of 26 individuals from 14 fancy breeds which were sampled within Synbreed project. The list of breeds with more details is presented in Table 1. OG breeds present a group of breeds that were not selected for commercial purpose such as egg or meat production. They consist of various breeds that were mainly selected for phonotypical traits such as feather color, feather style and comb style.

Table 1. Name, abbreviation, number of individuals and the egg color for each breed used in this study.

Breed	Abbreviation	# of lines	# of individuals	Egg color
White Leghorn	WL(1/2/3)	3	30(0♂,30♀)	White
Rhode Island Red	RIR(1/2)	2	20(2♂,18♀)	Brown
White Rock	WR(1/2)	2	20(2♂,18♀)	Brown
Asil	OG/Asil	1	2(0♂,2♀)	Brown
Brahma	OG/Brah	1	2(0♂,2♀)	Brown
Cochin	OG/Coch	1	2(0♂,2♀)	Brown
Fayoumi	OG/Fayo	1	2(0♂,2♀)	White
Gallus gallus gallus	OG/Ggal	1	2(0♂,2♀)	Brown
Gallus gallus spadiceus	OG/Gspa	1	2(0♂,2♀)	Brown
Green legged Partridge	OG/GreP	1	2(0♂,2♀)	White
Hungarian White Goedoelloe	OG/HunW	1	2(0♂,2♀)	Brown
Jaerhoens	OG/Jaer	1	2(0♂,2♀)	White
Malay	OG/Mala	1	2(0♂,2♀)	Brown
Marans	OG/Mara	1	2(0♂,2♀)	Brown
Orlov	OG/Orlo	1	2(0♂,2♀)	White
Paduaner	OG/Padu	1	1(0♂,1♀)	White
Transylvanian Naked Neck	OG/Tran	1	1(0♂,1♀)	Brown

DNA was isolated using a phenol/chloroform method for the DNA isolation [18] from whole blood collected from the wing vein using EDTA as anticoagulant. DNA quality and concentration of each sample was calculated and equal amounts of DNA were used for genotyping on three Affymetrix 600K SNP arrays using the Affymetrix® GeneTitan® system according to the procedure described by Affymetrix [19]. Data is available from the authors upon request.

This study was carried out in strict accordance with the German Animal Welfare regulations. The blood taking protocol was approved by the Committee of Animal Welfare at the Institute of Farm Animal Genetics of the Friedriech-Loeffler-Institut. Blood sampling was also notified to the Lower Saxonian authorities according to § 8a para. 1 of the German Animal Welfare Act. The blood takings were registered at the Lower Saxony State Office for Consumer Protection and Food Safety (Registration Number 33.9-42502-05-10A064).

Overlapping SNPs between the three 600K SNP arrays were removed and a total of 1,139,073 SNPs remained. To avoid imputation error in further analyses and due to the high amount of SNP and good coverage of the genome, 148,712 SNPs with at least one missing value were removed. Next the included SNPs were filtered for minor allele frequencies lower than 5% (74,202 were removed) in order to avoid genotyping errors, this approach was suggested by the data provider. The SNPs were located on autosomal chromosomes (1-28), one sex chromosome (Z), and two linkage groups, LGE22C19W28_E50C23 and LGE64, which were named Chr40 and Chr41, respectively. A total of 916,159 SNPs remained after filtering (throughout this paper, 916,159 is referred to as 1M SNPs). The entire filtering process was done by using the software PLINK (http://pngu.mgh.harvard.edu/purcell/plink/) [20].

Population structure analysis

Two methods were used in order to retrieve the structure of the studied samples; principal component analysis (PCA) using the R package ADEGENET [21,22] and maximum likelihood estimation of individual ancestries using ADMIXTURE software with several null hypotheses [23].

F_{ST} calculation and permutation test

To identify the regions under selection, Wright's F_{ST} [10] was calculated for all pairwise combinations of breeds and average F_{ST} values were calculated for overlapping windows along each chromosome. Each window consisted of 40 SNPs with an overlap of 20 SNPs with the next window. Average window size was 20,554 bp with a minimum of 2,029 bp and a maximum of 6,633,801 bp.

To assess distribution of the F_{ST} values we conducted a permutation test with 100 replications. For each replicate the individuals were randomly assigned to one of two groups, then F_{ST} was calculated for each SNP and averaged for the same windows as with the non-permuted data. The maximum and minimum F_{ST} value then was stored for each replicate.

Signatures of selection

According to the PCA and ADMIXTURE structural analysis (Figure 1 and Figure 2, respectively), breeds were arranged in six different groups; the two White Rock lines were pooled together (WR, n=20), each of the Rhode Island Red lines remained in one group (RIR1, n=10 and RIR2, n=10), White Leghorn line one was kept as one group (WL1, n=10), line two and line three from White Leghorn were pooled together (WL2&3, n=20), and all the non-commercial chicken breeds were pooled in one group (OG, n=26).

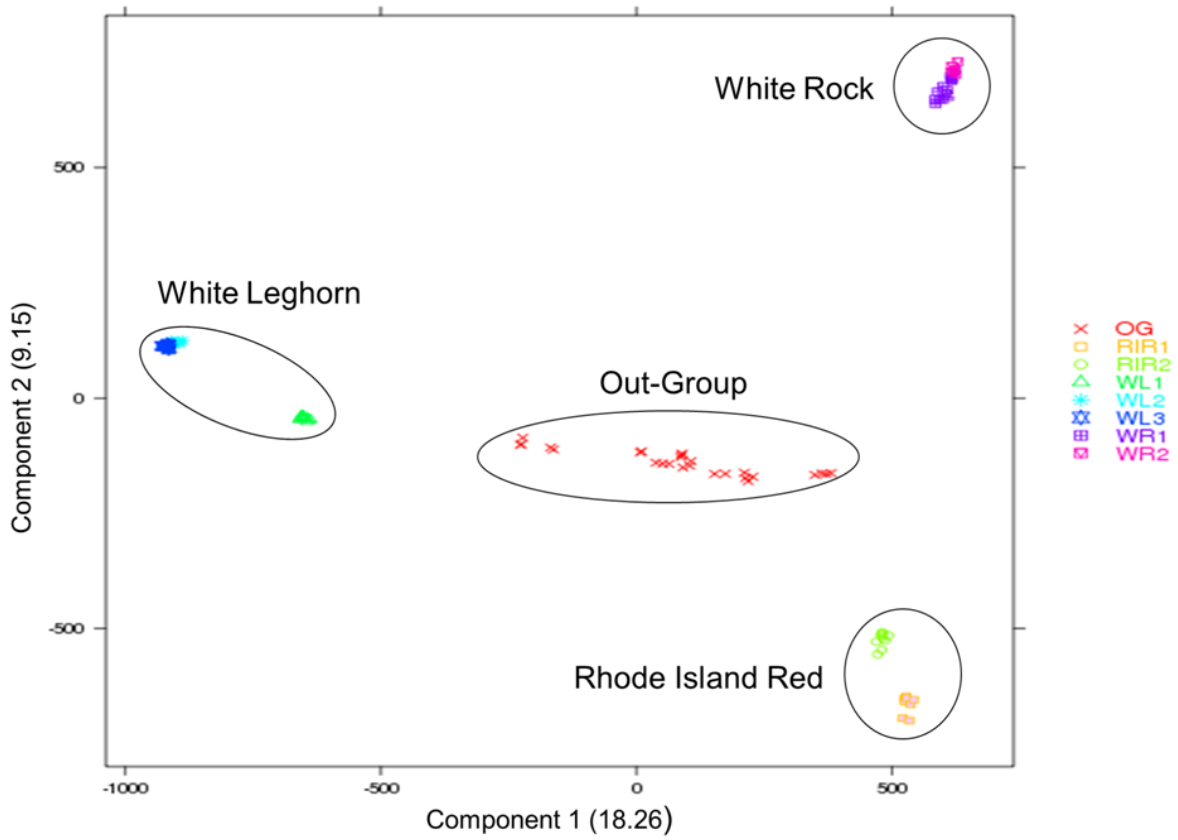

Figure 1. PCA analysis for all the 96 individuals with 1 million SNPs.

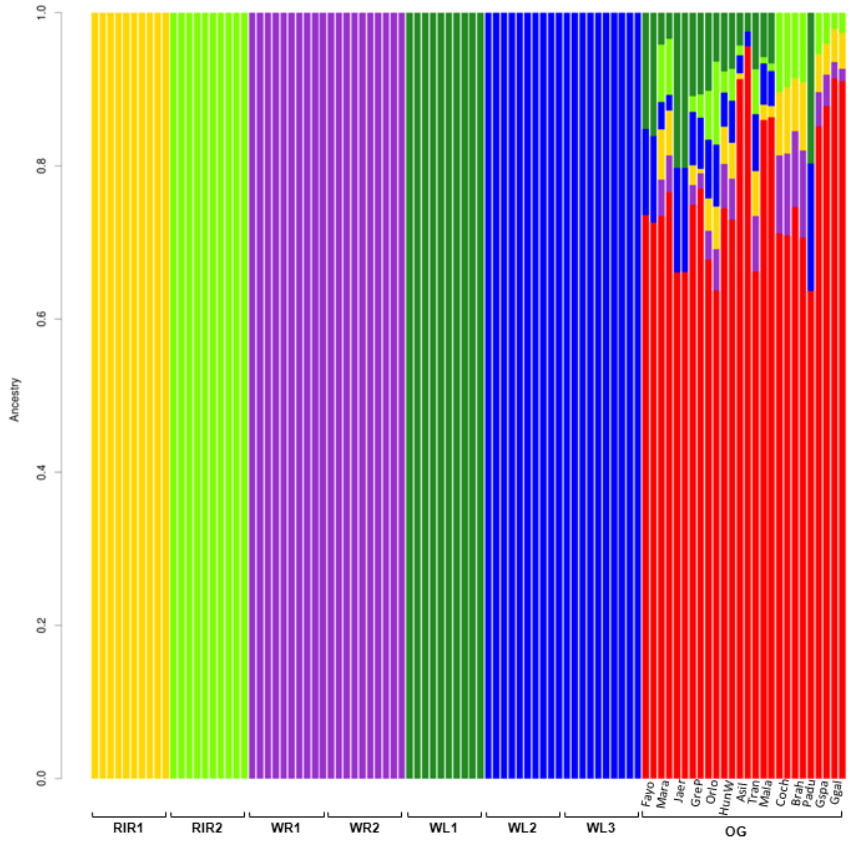

Figure 2. Result of ADMIXTURE structural analysis with null hypothesis of six breeds. Two rightmost individuals in OG are *Gallus gallus gallus*, and the third and fourth last individuals are *Gallus gallus spadiceus*.

Two sets of comparisons were made in this study in order to detect selection signatures. First, a comparison between commercial egg layers and the out-group (LY vs. OG) was carried out. For this comparison, F_{ST} values between the out-group and each of the commercial groups (RIR1, RIR2, WR, WL1 and WL2&3) were calculated for each SNP in the window and averaged. Second a comparison between white egg layers and brown egg layers (WL vs. BL) was conducted. In this case, the average of F_{ST} values between the white egg layers (WL1 and WL2&3) and the brown egg layers (RIR1, RIR2 and WR) in each window was calculated.

Next, based on the genome-wide distribution of F_{ST}, a threshold cutting of the upper and lower 1% was used for the definition of extreme values. To compensate for the higher average F_{ST} on sex-chromosome Z compared to the autosomes, the thresholds for chromosome Z were determined separately, by cutting of the upper and lower 1% of the F_{ST} distribution on chromosome Z [24].

Annotation

The regions with extreme F_{ST} values can be considered as good candidates for selective sweeps. For each comparison all the extreme windows (the upper or lower 1%) that were within 500 kb of each other were grouped to form a set of joined windows. For all joined windows gene annotation and pathway annotation was completed. Gene annotations were done with the biomaRt R package [25] based on Ensembl data base [26]. For pathway annotation KEGG database [27] was used. Fisher exact test was run for gene enrichment analysis for all annotated genes using DAVID (The Database for Annotation, Visualization and Integrated Discovery) [28,29]. We assumed pathways and gene ontologies with $p \leq 0.05$ as being under selection.

Results and Discussion

Components one and two of the PCA analysis with 1M SNPs, jointly accounting for 27.4 per cent of the total variance, are plotted in Figure 1. The commercial white egg-layer breeds were separated by component 1 from brown egg-layers. In addition, two brown egg-layer breeds (RIR and WR) were separated from each other by component 2. The outgroup is rather diverse and stays in the center of the distribution. As expected from the Lohmann breeding program, line two and line three of White Leghorns, and both lines in White Rock clustered together, respectively.

Additionally, based on the cross validation test of admixture with all the commercial breeds, maximum likelihood estimation of the individual ancestries under the null hypothesis of six populations was run for 1M SNPs. The result is shown in Figure 2. These analyses are largely in agreement with the expected historical origin of the breeds [6] and the result of the PCA. Admixture analysis clustered OG breeds as one group; however there was an admixture between different breeds in OG with layer breeds. Interestingly, there is no admixture between White Leghorns and ancestral chicken breeds (*Gallus gallus* and *Gallus spadiceus*).
Based on these results, individuals were arranged in six different breed groups of WL1, WL2&3, WR, RIR1, RIR2 and OG.

Average F_{ST} within brown layers (RIR vs. WR, 0.18) was lower than the average F_{ST} value between white layers and brown layers (RIR vs. WL (0.24) and WR vs. WL (0.26)) (shown in

Table 2), which shows that the similarity within the brown layers is higher than between white layers and brown layers, as it is expected. The average F_{ST} values along with the standard deviation for all group comparisons are shown in Table 3. In general, F_{ST} values between the out-group and commercial layer breeds are lower than the F_{ST} values between two commercial layer breeds, which is due to the fact that the allele frequency spectrum in commercial layers follows a U-shaped distribution while in the out-group it follows approximately a uniform distribution (results are not shown). F_{ST} values between lines of breeds are always lower than between breeds, which show the similarity within breeds is much higher than between breeds.

Table 2. Average F_{ST} values with standard deviation between different breeds.

	WL	**WR**
RIR	0.2419(±0.25)	0.1768(±0.20)
WR	0.2641(±0.27)	

Table 3. Average F_{ST} values with standard deviation over all SNPs for all compression.

	WL2and3	**RIR1**	**RIR2**	**WR**	**OG**
WL1	0.1543(±0.21)	0.2653(±0.31)	0.2524(±0.30)	0.2382(±0.29)	0.1184(±0.14)
WL2and3		0.2715(±0.32)	0.2590(±0.30)	0.2567(±0.30)	0.1570(±0.17)
RIR1			0.1148(±0.17)	0.1662(±0.23)	0.1006(±0.13)
RIR2				0.1523(±0.24)	0.0904(±0.11)
WR					0.1155(±0.13)

The permutation test showed that the F_{ST} distribution under randomization is much lower than the observed distribution of F_{ST} (results not shown). In all cases the minimum F_{ST} value obtained from the permuted data was close to zero and the maximum was around 0.3, which corresponds to a threshold 10 times lower than the threshold that we used, and is not helpful for the derivation of empirical threshold values. Based on F_{ST} values averaged in overlapping windows a total of 656 selective signatures (321 and 335 regions for the upper and lower 1% F_{ST} distribution, respectively) were detected when comparing commercial egg-layer breeds. In the comparison between non-commercial and commercial breeds, a total of 630 selective signatures (322 and 308 regions for the upper and lower 1% F_{ST} distribution respectively) were detected. The genome-wide distribution of F_{ST} values obtained with the comparison LY vs. OG and WL vs. BL are depicted in Figures 3 and 4, respectively.

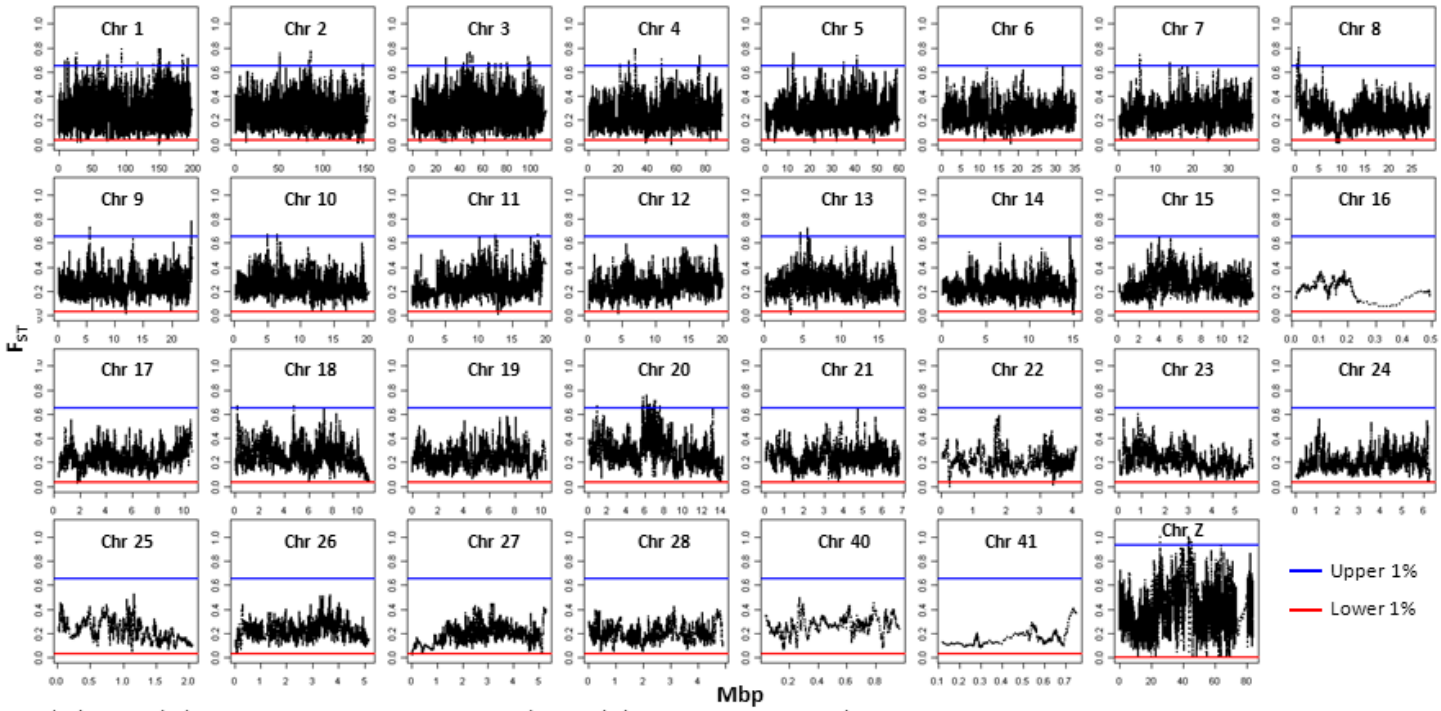

Figure 3. F_{ST}-values of overlapping windows for comparison between commercial layers and out-group. Red (blue) line indicates the upper (lower) 1% of F_{ST} distribution.

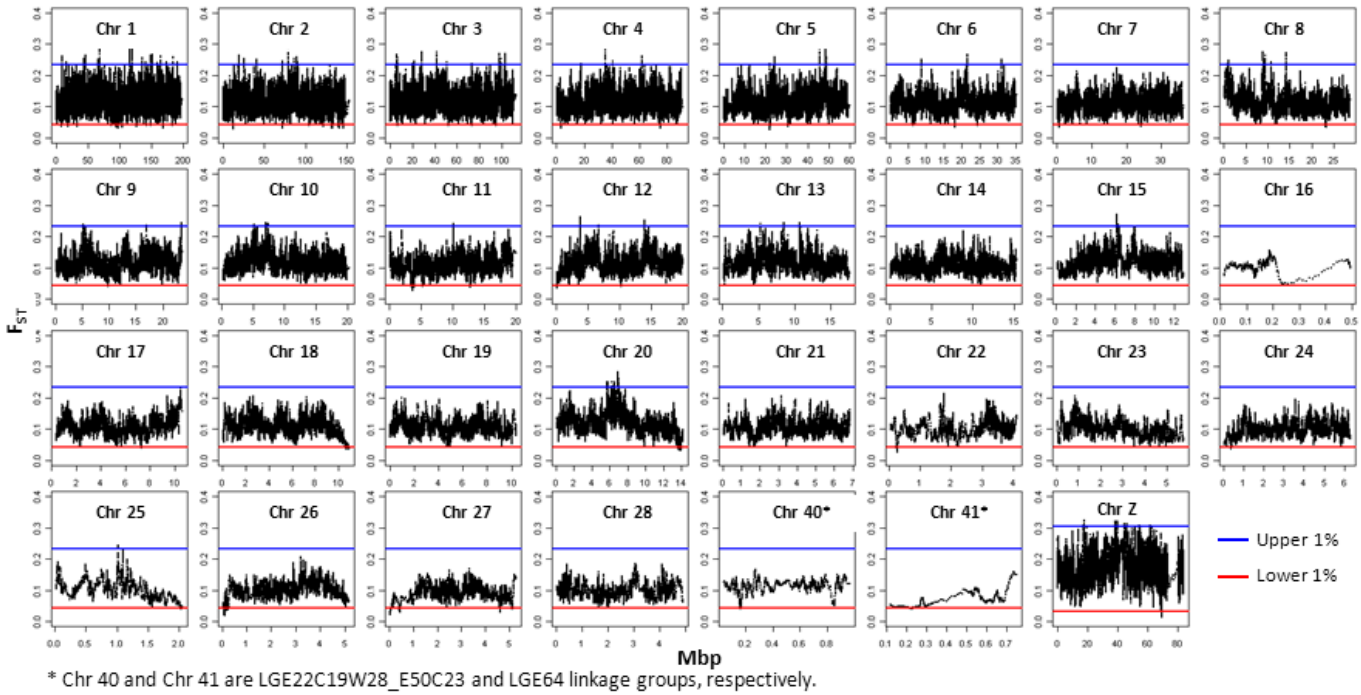

Figure 4. F_{ST}-values of overlapping windows for comparison between brown layers and white layers. Red (blue line) indicates the upper (lower) 1% of F_{ST} distribution.

The overlapping windows method was used for two reasons: to reduce the noisiness of single-locus statistics by combining data from several adjacent markers, and to avoid the risk of passing over genomic gaps. As Qanbari *et al.* [30] suggested, the use of overlapping windows has a higher power of detecting selective sweeps compared to sliding windows. In this work, defining a window size of 40 SNPs was a subjective decision, but it was motivated by previous studies [14,31] and the requirement of having sufficient coverage all over the genome. SNPs on each of the three Affymetrix 600K SNP arrays are distributed equally with respect to the genetic distance; this explains the large difference of window size based on bp. The outlier approach is an effective method for identifying the genes under selection lacking known phenotypes [32]. However, as Akey [33] explained, an outlier signal is not necessarily synonymous with regions being under selection.

Many of the detected outliers could be considered false positives. This might be the case because the F_{ST} calculations assume that the populations have the same effective size and were derived independently from the same ancestral population. The error caused by this assumption is similar to well-known effects of cryptic structure in genome-wide association studies [34].

Regions with F_{ST} values in the lower tail of the distribution are of interest for comparison of commercial breeds, which have been selected for very similar traits but starting from a very diverse genetic background, especially so for white and brown layers. In contrast, F_{ST} values in the upper tail of the distribution are of interest since they may display regions under selection for different breeding goals such as egg shell color. For comparing commercial breeds with the non-commercial breeds, the regions with F_{ST} values in the upper tail of the distribution are relevant because of the large contrast in breeding goals between these groups while the regions with F_{ST} values in the lower tail of the distribution might show regions that have been selected naturally or artificially before the intense selection on laying performance in commercial breeds.

Annotation was carried out for all regions with extreme F_{ST} values, i.e. potential selective sweeps. The lists of genes for selective sweeps are available in the supplementary tables (Table S1, S2, S3 and S4). In general, the annotation list is enriched with genes of biological interest involved in various pathways such as cellular amino acid catabolic process (p= 0.012), regulation of growth (p= 0.012), calcium ion binding (p= 0.033), B cell activation (p= 0.031), immune system development (p= 0.034) and post-embryonic development (p= 0.035), all of which could be related to production traits indirectly. The lists of pathways and gene ontologies under selection are available in the supplementary tables (Table S5, S6, S7 and S8). In both comparisons (LY vs. OG and WL vs. BL), we were able to identify several genes related to the breeding goals of egg-layer chickens, such as the age at sexual maturity, laying rate, body weight, and feed conversion [35] (Table 4).

Table 4. Genes associated to productive traits in both comparisons. ≠ symbol stands for difference between two group and = symbol stand for similarity between two groups. B and W stand for comparison between brown and white egg layers and L and G stand for comparison between commercial layers and out-group.

Gene	Chr	Function	Comparison
SREBF2	1	Involved in the rapid growth stages of follicle development.	B≠W
POU1F1	1	Associated with growth performance in chicken.	L=G
CST3	3	Involve in calcium release into the medium.	B=W
TGFB2	3	Significantly associated with chicken growth traits and not associated with any reproduction traits	L≠G,B≠W
CAPN1	3	Associated with meat quality traits in chicken.	B≠W
NCOA1	3	Associated with total egg production at (age 300 day) and age at first egg	L≠G
POMC	3	Associated with feed conversion and body weight in commercial broiler	L≠G
SPP1	4	Associated with 5-week body weight and quality of egg shells in laying hens	L≠G
IGFII	5	Influencing growth and carcass traits.	B≠W
RALGAPA1	5	Associated with reproductive traits and broodiness.	B≠W
IGFBP2	7	Associated with body composition, body weight, and affects fatness traits in chickens	L≠G , B=W
PRKAB2	8	Associated with live-weight, carcass-weight, leg-muscle-weight and abdomen-fat-weight	L≠G
CCT6A	19	Associated with sexual maturity in hens.	L=G
IL 19	26	Assoiciated with responses to intracellular poultry pathogens like bacteria and protozoa.	L=G
AMH	28	Expression is significantly greater in broiler breeder hens as compared with laying hens.	L=G
SLC45A2	Z	Inhibitor of expression of red pheomelanin in Silver chickens.	L=G

Many genes were identified in selective sweep regions in the comparison between brown and white layers. *TGFb2*, *CAPN1* and *IGF2* were all located in regions that were different between brown and white layers. *TGFb2* (transforming growth factor, beta 2) is significantly associated with chicken growth traits and is not associated with any reproduction traits [36]. *TGFb2* is expressed 4-fold greater in broiler compared with layer hens at 15 weeks of age [37]. *CAPN1* is associated with meat quality [38,39]. *IGF2* (insulin-like growth factor 2), which is believed to be a major fetal growth factor in contrast to insulin-like growth factor 1 [40], has a great influence on growth and carcass traits in chicken [41]. The presence of genes associated with meat quality and production in regions that were different between brown and white layers reflects the fact that brown egg-layers were originally a dual-purpose breed. Specifically, brown layers were bred for meat production as well as egg-production, whereas white egg layers were bred only for egg production [6]. *SREBF2* and *RALGAPA1*, which are both associated with reproductive traits and broodiness [42,43], were also located in the regions with high contrast between the two layer breeds. This can indicate that different regions were selected for reproductive traits in the different egg-layer breeds.

In the comparison of commercial-layers and out-group, *NCOA1*, which corresponds to the total egg production (at age 300 days) and age at first egg [44], along with *SPP1*, which is associated with 5-week body weight and quality of egg shells in laying hens [45], were

located in the regions that were different between commercial-layers and out-group chicken. This may reflect the intense selection of the regions associated with egg production and quality traits in laying breeds. *PRKAB2*, *POMC* and *TGFb2* which are associated with live-weight, carcass-weight, leg-muscle-weight, abdomen-fat-weight and feed conversion, were also located in the regions that differ between commercial-layers and out-group chicken [36,46,47]. This may be due to the existence of brown egg layers with a dual purpose ancestral background in the commercial-layers group under study.

IGFBP2, which inhibits or stimulates the growth promoting effects of the *IGFs* [48],is associated with body composition, body weight and affects fatness traits in chickens [49,50], was identified both in the similarity between the two layer breeds (white layers and brown layers) and in the difference between the layers and the out-group. This indicates positive selection of this gene in both groups of layer breeds, although they have different genetic background and they have been selected separately.

Several further regions were identified as selective signatures in the comparison between commercial lines and the out-group. These regions mainly corresponded to primary genes such as *CCT6A* and *IL19*. *CCT6A* is a gene associated with sexual maturity in hens [51], and *IL19* plays an important role in responses to intracellular poultry pathogens like bacteria and protozoa [52]. *POU1F1* and *AMH*, which are both genes related to the growth performance in broiler chickens [53–55], were identified in regions that show similarity between the out-group and commercial layers.

In this study, we have identified more regions as putative selective sweeps compared to previously reported data by Rubin *et al.* [14] and Elferink *et al.* [16]. However, several of these regions were not associated with any genes related to production traits. This could be due to insufficient knowledge about these regions or it could also reflect false positives caused by genetic drift following the separation of the breeds [56]. Although, we have not annotated selection signatures reported in other studies [14,16], our results agree with previously reported findings with respect to identified homologs of the same genes. For instance, *IGF2* is a homolog of *IGF1* which was identified in two studies [14,16]. We also identified *POU1F1* which binds to and transactivates promoters of growth hormone (GH) and thyroid-stimulating hormone chain (*TSHB*)-encoding genes [57], which were identified by Rubin and Elferink [14,16]. Another reason that the detected genes are different from the previous studies [14,16] could be that our study was based on layer breeds while other works included broiler breeds.

In conclusion we were able to identify several putative selective signature regions with genes corresponding to the productions traits layer breeds were selected for. These identified regions are good candidates for further studies. It was demonstrated that layers with a specific breeding history, which has led to animals with a very similar performance profile coming from a much differentiated genetic background, provide a unique opportunity for a better understanding of farm animal selection.

Supporting Information

Supporting information is available online.
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0094509

Reference

1. Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet 10: 639–650. doi:10.1038/nrg2611.

2. Darwin C (1883) The Variation of Animals and Plants Under Domestication. Macmillan publishers limited.

3. Kanginakudru S, Metta M, Jakati RD, Nagaraju J (2008) Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. BMC Evol Biol 8: 174. doi:10.1186/1471-2148-8-174.

4. Liu Y-P, Wu G-S, Yao Y-G, Miao Y-W, Luikart G, et al. (2006) Multiple maternal origins of chickens: Out of the Asian jungles. Mol Phylogenet Evol 38: 12–19. doi:10.1016/j.ympev.2005.09.014.

5. Al-Nasser A, Al-KHALAIFA H, Al-Saffar A, Khalil F, Albahouh M, et al. (2007) Overview of chicken taxonomy and domestication. Worlds Poult Sci J 63: 285–300. doi:10.1017/S004393390700147X.

6. Crawford R (1990) poultry breeding and genetics. Elsevier science.

7. Burt DW (2005) Chicken genome: Current status and future opportunities. Genome Res 15: 1692–1698. doi:10.1101/gr.4141805.

8. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23: 23–35. doi:10.1017/S0016672300014634.

9. Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet 10: 639–650. doi:10.1038/nrg2611.

10. Wright S (1949) The Genetical Structure of Populations. Ann Eugen 15: 323–354. doi:10.1111/j.1469-1809.1949.tb02451.x.

11. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. Nat Genet 40: 340–345. doi:10.1038/ng.78.

12. Myles S, Tang K, Somel M, Green RE, Kelso J, et al. (2007) Identification and Analysis of Genomic Regions with Large Between-Population Differentiation in Humans. Ann Hum Genet 0: 070810012647003–??? doi:10.1111/j.1469-1809.2007.00390.x.

13. Brown WRA, Hubbard SJ, Tickle C, Wilson SA (2003) The chicken as a model for large-scale analysis of vertebrate gene function. Nat Rev Genet 4: 87–98. doi:10.1038/nrg998.

14. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, et al. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464: 587–591. doi:10.1038/nature08832.

15. Johansson AM, Pettersson ME, Siegel PB, Carlborg Ö (2010) Genome-Wide Effects of Long-Term Divergent Selection. PLoS Genet 6: e1001188. doi:10.1371/journal.pgen.1001188.

16. Elferink MG, Megens H-J, Vereijken A, Hu X, Crooijmans RPMA, et al. (2012) Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds. PLoS ONE 7: e32720. doi:10.1371/journal.pone.0032720.

17. Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, et al. (2013) Development of a high density 600K SNP genotyping array for chicken. BMC Genomics 14: 59. doi:10.1186/1471-2164-14-59.

18. Green MR, Sambrook J (2012) Molecular cloning: a laboratory manual. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

19. Axiom Genotyping Solution Data Analysis Guide (2011).

20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81: 559–575.

21. Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27: 3070–3071. doi:10.1093/bioinformatics/btr521.

22. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405. doi:10.1093/bioinformatics/btn129.

23. Zhou H, Alexander D, Lange K (2011) A quasi-Newton acceleration for high-dimensional optimization algorithms. Stat Comput 21: 261–273. doi:10.1007/s11222-009-9166-3.

24. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12: 1805–1814. doi:10.1101/gr.631202.

25. Durinck S (n.d.) biomaRt: Interface to BioMart databases (e.g. Ensembl, COSMIC ,Wormbase and Gramene). R package version 260.

26. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, et al. (2012) Ensembl 2013. Nucleic Acids Res 41: D48–D55. doi:10.1093/nar/gks1236.

27. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40: D109–114. doi:10.1093/nar/gkr988.

28. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57. doi:10.1038/nprot.2008.211.

29. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1–13. doi:10.1093/nar/gkn923.

30. Qanbari S, Strom TM, Haberer G, Weigend S, Gheyas AA, et al. (2012) A High Resolution Genome-Wide Scan for Significant Selective Sweeps: An Application to

Pooled Sequence Data in Laying Chickens. PLoS ONE 7: e49525. doi:10.1371/journal.pone.0049525.

31. Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, et al. (2011) Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics 12: 318. doi:10.1186/1471-2164-12-318.

32. Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. Mol Ecol Resour 11: 184–194. doi:10.1111/j.1755-0998.2011.02987.x.

33. Akey JM (2009) Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res 19: 711–722. doi:10.1101/gr.086652.108.

34. Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. Nat Rev Genet 11: 459–463. doi:10.1038/nrg2813.

35. Muir WM, Aggrey SE (2003) Poultry Genetics, Breeding and Biotechnology. CABI. 724 p.

36. Tang S, Ou J, Sun D, Zhang Y, Xu G, et al. (2011) A novel 62-bp indel mutation in the promoter region of transforming growth factor-beta 2 (TGFB2) gene is associated with body weight in chickens. Anim Genet 42: 108–112. doi:10.1111/j.1365-2052.2010.02060.x.

37. Bennett AK, Hester PY, Spurlock DM (2007) Relationships of a transforming growth factor-beta2 single nucleotide polymorphism and messenger ribonucleic acid abundance with bone and production traits in chickens. Poult Sci 86: 829–834.

38. Zhang Z-R, Zhu Q, Jiang X-S, Du H-R (2007) [Study on correlation between single nucleotide polymorphism of CAPN1 gene and muscle tenderness and carcass traits in chicken]. Yi Chuan Hered Zhongguo Yi Chuan Xue Hui Bian Ji 29: 982–988.

39. Zhang ZR, Liu YP, Jiang X, Du HR, Zhu Q (2008) Study on association of single nucleotide polymorphism of CAPN1 gene with muscle fibre and carcass traits in quality chicken populations. J Anim Breed Genet Z Für Tierz Zücht 125: 258–264.

40. Jiao S, Ren H, Li Y, Zhou J, Duan C, et al. (2013) Differential regulation of IGF-I and IGF-II gene expression in skeletal muscle cells. Mol Cell Biochem 373: 107–113. doi:10.1007/s11010-012-1479-4.

41. Tang S, Sun D, Ou J, Zhang Y, Xu G, et al. (2010) Evaluation of the IGFs (IGF1 and IGF2) genes as candidates for growth, body measurement, carcass, and reproduction traits in Beijing You and Silkie chickens. Anim Biotechnol 21: 104–113. doi:10.1080/10495390903328090.

42. Shen X, Zeng H, Xie L, He J, Li J, et al. (2012) The GTPase activating Rap/RanGAP domain-like 1 gene is associated with chicken reproductive traits. PloS One 7: e33851. doi:10.1371/journal.pone.0033851.

43. Seol HS, Sato K, Matsubara Y, Schneider WJ, Akiba Y (2007) Modulation of sterol regulatory element binding protein-2 in response to rapid follicle development in chickens. Comp Biochem Physiol B Biochem Mol Biol 147: 698–703. doi:10.1016/j.cbpb.2007.04.012.

44. Huang HY, Li SF, Zhao ZH, Liang Z, Zhang J, et al. (2011) Association of polymorphisms for nuclear receptor coactivator 1 gene with egg production traits in the maternal line of Shaobo hens. Br Poult Sci 52: 328–332. doi:10.1080/00071668.2011.577057.

45. Arazi H, Yoselewitz I, Malka Y, Kelner Y, Genin O, et al. (2009) Osteopontin and calbindin gene expression in the eggshell gland as related to eggshell abnormalities. Poult Sci 88: 647–653. doi:10.3382/ps.2008-00387.

46. Wang J-L, Zhu Q, Liu Y-P, DU H-R (2008) [Associations between SNP of chicken PRKAB2 gene and slaughter and meat quality traits]. Yi Chuan Hered Zhongguo Yi Chuan Xue Hui Bian Ji 30: 1033–1038.

47. Sharma P, Bottje W, Okimoto R (2008) Polymorphisms in uncoupling protein, melanocortin 3 receptor, melanocortin 4 receptor, and pro-opiomelanocortin genes and association with production traits in a commercial broiler line. Poult Sci 87: 2073–2086. doi:10.3382/ps.2008-00060.

48. Van Buul-Offers SC, Van Kleffens M, Koster JG, Lindenbergh-Kortleve DJ, Gresnigt MG, et al. (2000) Human insulin-like growth factor (IGF) binding protein-1 inhibits IGF-I-stimulated body growth but stimulates growth of the kidney in snell dwarf mice. Endocrinology 141: 1493–1499.

49. Li ZH, Li H, Zhang H, Wang SZ, Wang QG, et al. (2006) Identification of a single nucleotide polymorphism of the insulin-like growth factor binding protein 2 gene and its association with growth and body composition traits in the chicken. J Anim Sci 84: 2902–2906. doi:10.2527/jas.2006-144.

50. Leng L, Wang S, Li Z, Wang Q, Li H (2009) A polymorphism in the 3'-flanking region of insulin-like growth factor binding protein 2 gene associated with abdominal fat in chickens. Poult Sci 88: 938–942. doi:10.3382/ps.2008-00453.

51. Kang L, Zhang Y, Zhang N, Zang L, Wang M, et al. (2012) Identification of differentially expressed genes in ovaries of chicken attaining sexual maturity at different ages. Mol Biol Rep 39: 3037–3045. doi:10.1007/s11033-011-1066-x.

52. Kim S, Miska KB, McElroy AP, Jenkins MC, Fetterer RH, et al. (2009) Molecular cloning and functional characterization of avian interleukin-19. Mol Immunol 47: 476–484. doi:10.1016/j.molimm.2009.08.027.

53. Tao Y, Li G-H, Hu Y-P, Mekki MD, Chen K-W, et al. (2008) [Genetic effects of MC4R and POU1F1 gene on the growth performance in Jinghai yellow chicken]. Yi Chuan Hered Zhongguo Yi Chuan Xue Hui Bian Ji 30: 900–906.

54. Nie Q, Fang M, Xie L, Zhou M, Liang Z, et al. (2008) The PIT1 gene polymorphisms were associated with chicken growth traits. BMC Genet 9: 20. doi:10.1186/1471-2156-9-20.

55. Johnson PA, Kent TR, Urick ME, Trevino LS, Giles JR (2009) Expression of anti-Mullerian hormone in hens selected for different ovulation rates. Reprod Camb Engl 137: 857–863. doi:10.1530/REP-08-0406.

56. Der R, Epstein CL, Plotkin JB (2011) Generalized population models and the nature of genetic drift. Theor Popul Biol 80: 80–99. doi:10.1016/j.tpb.2011.06.004.

57. Jiang R, Li J, Qu L, Li H, Yang N (2004) A new single nucleotide polymorphism in the chicken pituitary-specific transcription factor (POU1F1) gene associated with growth rate. Anim Genet 35: 344–346. doi:10.1111/j.1365-2052.2004.01164.x.

3rd Chapter

Genome scan for selection in structured layer chicken populations exploiting linkage disequilibrium information

Mahmood Gholami[1], Christian Reimer[1], Malena Erbe[1], Rudolf Preisinger[2], Annett Weigend[3], Steffen Weigend[3], Bertrand Servin[4] and Henner Simianer[1]

1 Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Göttingen, Albrecht-Thaer-Weg 3, 37075 Göttingen
2 LOHMANN Tierzucht GMBH, Am Seedeich 9-11, P.O.Box 460, 27454 Cuxhaven
3 Institute of Farm Animal Genetics (ING), Friedrich-Loeffler-Institut (FLI), Höltystraße 10, 31535 Neustadt
4 Laboratoire Génétique, Physiologie et Systèmes d'Elevage , Institut National de la Recherche Agronomique, 31320 Castanet-Tolosan, France

Submitted to PLOS ONE

Abstract

An increasing interest is being placed in the detection of genes, or genomic regions, that have been targeted by selection because identifying signatures of selection can lead to a better understanding of genotype-phenotype relationships. A common strategy for the detection of selection signatures is to compare samples from distinct populations and to search for genomic regions with outstanding genetic differentiation. The aim of this study was to detect selective signatures in layer chicken populations using a recently proposed approach, hapFLK, which exploits linkage disequilibrium information while accounting appropriately for the hierarchical structure of populations. We performed the analysis on 70 individuals from three commercial layer breeds (White leghorn, White Rock and Rhode Island Red), genotyped for approximately 1 million SNPs. We found a total of 41 and 107 regions with outstanding differentiation or similarity using hapFLK and its single SNP counterpart FLK respectively. Annotation of selection signature regions revealed various genes and QTL corresponding to productions traits, for which layer breeds were selected. A number of the detected genes were associated with growth and carcass traits, including *IGF-1R*, *AGRP* and *STAT5B*. We also annotated an interesting gene associated with the dark brown feather color mutational phenotype in chickens (*SOX10*). We compared F_{ST}, FLK and hapFLK and demonstrated that exploiting linkage disequilibrium information and accounting for hierarchical population structure improved the power of detection in our data set.

Introduction

A local reduction of genetic variation, commonly referred to as a "selective sweep", is caused by the rapid fixation of a beneficial mutation [1]. Study such signatures of selection can provide valuable insights into genomic regions harboring interesting genes that are or have been under selective pressure and hence can help to understand the mechanisms that led to the differentiation of various genotypes and their influenced phenotypes during selection. Recently, an increasing interest has been placed in the detection of genes, or genomic regions, that are targeted by selection [2], permitted by the availability of large-scale SNP datasets that allow to scan the genome for positions that may have been targets of recent selection [3].

Many different methods are available for detecting selective sweeps from DNA sequence data. Qanbari *et al.* (2014) [4] classified these methods in two main groups: intra-population statistics (e.g. Kim and Nielsen (2004) [5] and Sabeti *et al.* (2002) [3]) and inter-populations

statistics (e.g. Lewontin and Krakauer (1973) [6] and Beaumont and Balding (2004) [7]). Innan and Kim (2008) [8] and Yi *et al.* (2010) [9] showed that between recently diverged populations, inter-populations statistics have more statistical power for the detection of selection signatures. These methods are particularly suited for studying species that are structured in well-defined populations, which is the case in many domesticated species.

Inter-populations statistics can be divided into two groups based on single site or haplotype differentiation analyses [4]. The most widely used single site differentiation statistic is Wright's fixation index, F_{ST} [10]. A major concern with Wright's F_{ST} is that it implicitly assumes that populations have the same effective size (N_e) and to be derived independently from an ancestral population. When this is not true F_{ST} will produce false positive signals, similar to the well-known effects of cryptic structure in genome-wide association studies [11]. Bonhomme *et al.* (2010) [12] proposed a new statistic, termed FLK, that deals with N_e variation and historical branching of populations by incorporating a population kinship matrix into the Lewontin and Krakauer (LK) [6] statistic and showed that FLK is indeed more powerful than F_{ST} for a given false positive rate.

Another group of methods builds on the fact that haplotype diversity and linkage disequilibrium (LD) patterns contain useful information for the detection of selection signatures [13] and therefore, usage of haplotype or LD based differentiation analyses has its own advantages. Browning and Weir (2010) [14] showed that SNP ascertainment bias has less impact on haplotype based differentiation analyses compared to single site differentiation. A major challenge regarding the haplotype differentiation scans is that it does not account for the possibility of hierarchical structure between populations. Therefore Fariello *et al.* (2013) [15] proposed the hapFLK statistic which is a haplotype based extension of the FLK statistic that accounts for both hierarchical structure and haplotype information. They showed that using haplotype information to detect selection in F_{ST}-like approaches greatly increases the detection power. Specifically, they demonstrated that the hapFLK statistic has more power in detecting soft sweeps, incomplete sweeps and sweeps occurring in several populations.

The growing genomic resources, the relatively rapid reproduction time and the existence of several inbred lines together with strong agricultural interest make chicken an excellent model for studying the signatures of selection under artificial breeding conditions [16]. Several studies have investigated selection signatures in chicken either using sequence data or genotype data from low to medium density SNP chips. Rubin *et al.* (2010) [17] studied the signatures of domestication and selective sweeps using the "Pooled Heterozygosity" (H_P)

statistic in various commercial broiler and layer lines. Johansson *et al.* (2010) [18] explored the genomes of two lines of chickens subjected to 50 generations of divergent selection using a 60k SNP assay. Qanbari *et al.* (2012) [19] applied a modified sliding window, called "creeping window", of H_P measures in pooled sequence data in laying chickens. In an earlier work we [20] studied the signatures of selection by F_{ST} in seven commercial breeds using approximately one million SNPs which, however, ignored the hierarchical structure of the populations analyzed. Recent divergence of certain commercial breeds [21] and the introduction of strong selection for production traits (in the 20th century) [22] fosters the interest in detecting selective sweeps in chicken using statistical methods that account for the strong hierarchical structure between these populations. Therefore, this dataset offers an interesting opportunity to evaluate methods that account for population structure in a setting characterized by a strong past selection pressure, high genetic drift and clear population structure, which has never been done before.

In this study, FLK [12] and hapFLK [15] statistics were applied on the same data as in our previous study on selection signatures in commercial chicken [20], allowing a comparison between F_{ST}, FLK and hapFLK. In contrast to our previous work, the approaches used in the current study have the potential to identify genomic regions which have been selected more recently (e.g. soft sweeps) and are associated with specific layer traits.

Materials and methods

Animals, Data collection and filtering

Two sets of samples - commercial egg layers and ancestral breeds (coded respectively LAY and ANC) - were used in this study. The commercial individuals from Lohmann Tierzucht GmbH originated from three different breeds. One commercial white egg layer breed based on White Leghorn (WL), with three separate lines, and the other two brown egg layer breeds based on White Rock (WR) and Rhode Island Red (RIR), respectively, each with two separate lines per breed. In each of these seven lines, ten individuals were sampled and genotyped. The ancestral breeds, comprising Red Jungle fowl (Cochin-Chinese) (*G. g. gallus*) and Red Jungle fowl (Burmese) (*G. g. spadiceus*) were sampled within the AVIANDIV project. A more detailed list of breeds is presented in Table 1. The ANC group consisted of two breed that are believed to be ancestors of modern chicken which have not been under selection for commercial purpose.

Table 1. Name, abbreviation, number of individuals and the egg color for each breed used in this study.

Breed	Abbreviation	# of lines	# of individuals	Egg color
White Leghorn	WL(1/2/3)	3	30(0♂,30♀)	White
Rhode Island Red	RIR(1/2)	2	20(2♂,18♀)	Brown
White Rock	WR(1/2)	2	20(2♂,18♀)	Brown
Gallus gallus gallus	ANC/Ggal	1	2(0♂,2♀)	Brown
Gallus gallus spadiceus	ANC/Gspa	1	2(0♂,2♀)	Brown

Genotyping was done with three Affymetrix 600K SNP arrays. Overlapping SNPs between the three 600K SNP arrays were removed by the data provider and a total of 1,139,073 SNPs remained. For this study we included only the SNPs that were located on autosomal chromosomes (1-28), SNPs that were located on sex chromosomes and linkage groups were removed (62,337 were removed). SNPs with at least one missing value and SNPs with minor allele frequencies lower than 5% (172,344 SNPs) were removed in order to avoid dealing with genotyping errors; this approach was suggested by the data provider. A total of 904,392 SNPs remained after filtering. The entire filtering process was done using the PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/) [23].

Population structure analysis

Two methods were used in order to retrieve the structure of the studied samples: construction of the phylogenetic tree using Reynolds' genetic distances [24], and maximum likelihood estimation of individual ancestries using ADMIXTURE software with several number of sub-populations [25].

FLK and hapFLK calculation

To identify regions under selection, FLK and hapFLK were calculated in all LAY breeds, using ANC individuals for rooting the population tree. FLK calculates variation of the inbreeding coefficient and incorporate hierarchical structure by using a population kinship matrix (for details see Bonhomme *et al.* (2010) [12]). The same matrix is used in hapFLK, but the statistic is computed from haplotype frequencies rather than SNP allele frequencies. Here, the haplotypes considered are in fact latent states extracted from the multipoint linkage

disequilibrium model of Scheet and Stephens [26] (for details read Fariello *et al.* (2013) [15]). To determine the number of underlying latent states we used the fastPHASE [26] cross validation procedure, which indicated that 5 or 10 haplotype clusters were adequate. We found that using either 5 or 10 haplotype clusters gave nearly identical results and therefore present those obtained assuming 5 haplotype clusters.

Assigning signatures of selection to specific population groups.

When using differentiation-based approaches, it is sometimes difficult to pinpoint the population(s) that have been the target of selection. Fariello *et al.* (2013) [15] proposed to decompose the hapFLK statistic by projecting it on principal components (PC) of the population kinship matrix to identify which part of the population tree exhibits an outlying differentiation in a particular genomic region. Here, we employed this approach to look for selection signatures that affected either (i) the whole population set (LAY), (ii) white layer populations or (iii) brown layer populations. For (i) we used the hapFLK statistic, for (ii) and (iii) we considered the projection of the statistic on the subtree corresponding to white (resp. brown) layer populations. In each case we considered that a position lying in the top or bottom 0.05% of the empirical distribution was potentially within a selection signature.
For each selection signature, we then re-estimated the branch lengths of the population tree, using local allele or haplotype clusters frequencies (see Fariello *et al.* (2013) [15] for details) and identified the branch lengths that seem significantly larger than the branches of whole genome tree to pinpoint selected populations.

Fitting of gamma distribution

As hapFLK statistic does not follow a known distribution under neutrality, the null distribution has to be estimated from the data. As hapFLK is similar to FLK, a good approximation to the asymptotic distribution of hapFLK comes from the gamma distribution family. To estimate p-values of selection signatures, we fitted a gamma distribution to the hapFLK observed distribution, using the minimum distance estimation method [25] which is robust to outliers, which helps to reduce the influence of selection signatures in estimating the null distribution. This was done for false detection rate (FDR) estimation.

Annotation

As explained above, regions with extreme FLK and hapFLK values were considered as candidates for selective sweeps. For all the three groups (all layers, white layers and brown layers) the extreme values (the upper or lower 0.05%) that were within 500 kb of each other were grouped together. For all joined groups gene annotations, QTL annotations and pathway annotations were completed. Gene annotations were done with the biomaRt R package [28] based on the Ensembl database [29]. Animal QTL database [30] was used for QTL annotation, KEGG database for pathway annotation [31] and Gene Ontology (GO) database for GO annotation [32]. Gene enrichment analysis was done with Fisher's exact test [33] for all annotated genes in all groups (all layers, white layers and brown layers) separately. Pathways and gene ontologies with $p \leq 0.05$ were identified as being under selection.

Results

Population structure

Based on the cross validation test of admixture with all the commercial breeds, maximum likelihood estimation of the individual ancestries assuming five underlying populations was run for 1M SNPs. The result is shown in Figure 1. A phylogenetic tree based on Reynolds' genetic distances with 100,000 randomly selected SNPs was constructed and is shown in Figure 2. Both analyses show basically the same result: the commercial white egg-layer breeds were separated from brown egg-layers and grouped in one sub-tree. In the sub-trees, the two white-layer lines WL2 and WL3 as well as the two brown-layer lines WR1 and WR2 form a separate sub-cluster, respectively. The population specific fixation indices of all populations, also shown in Figure 2, are extremely high (ranging from 0.45 to 0.75), reflecting the very strong effect of genetic drift in these populations, with the three White Leghorn populations notably more inbred than the Brown layer populations.

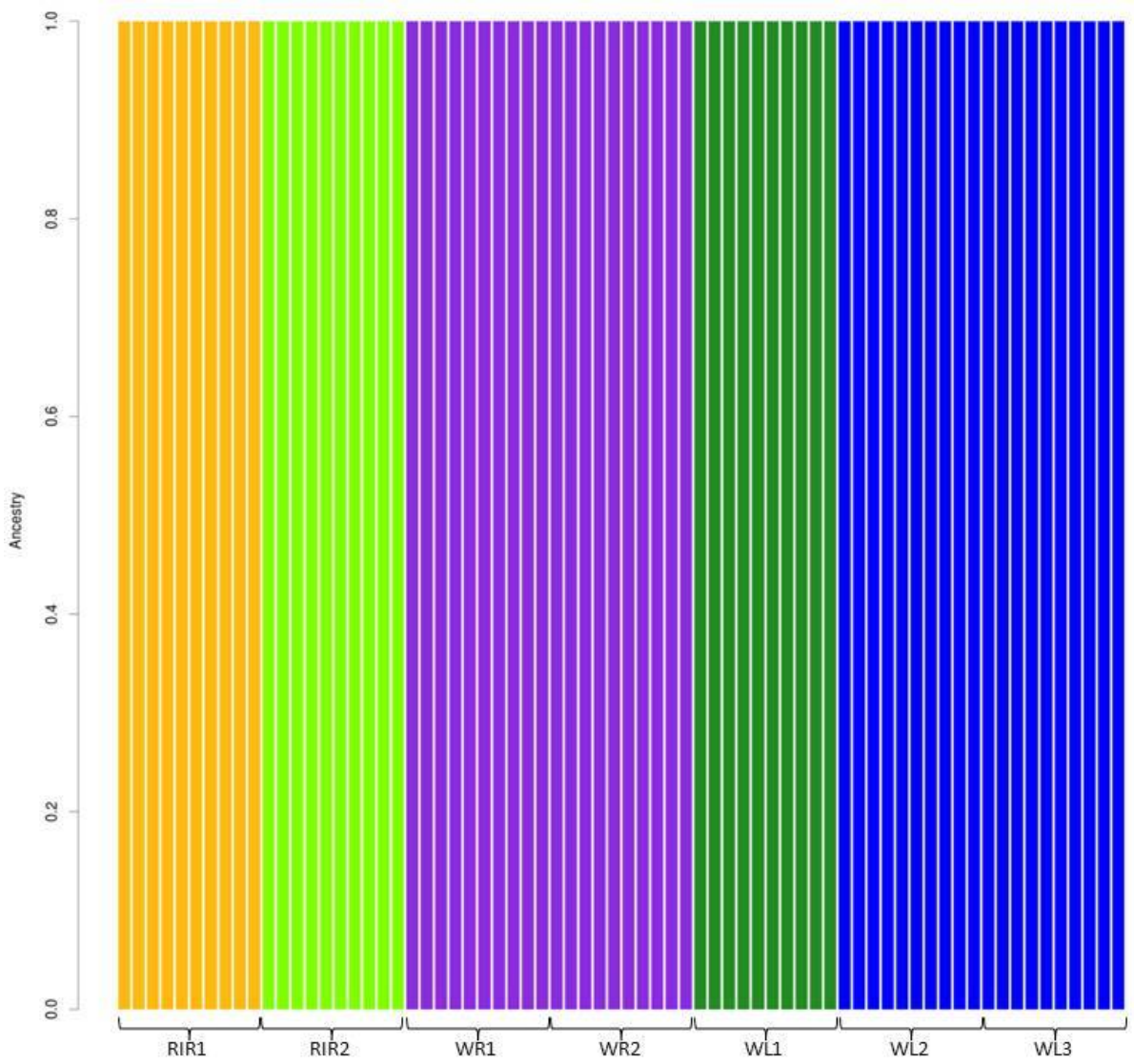

Figure 1. Maximum likelihood estimation of individual ancestries using ADMIXTURE with null hypothesis of five populations.

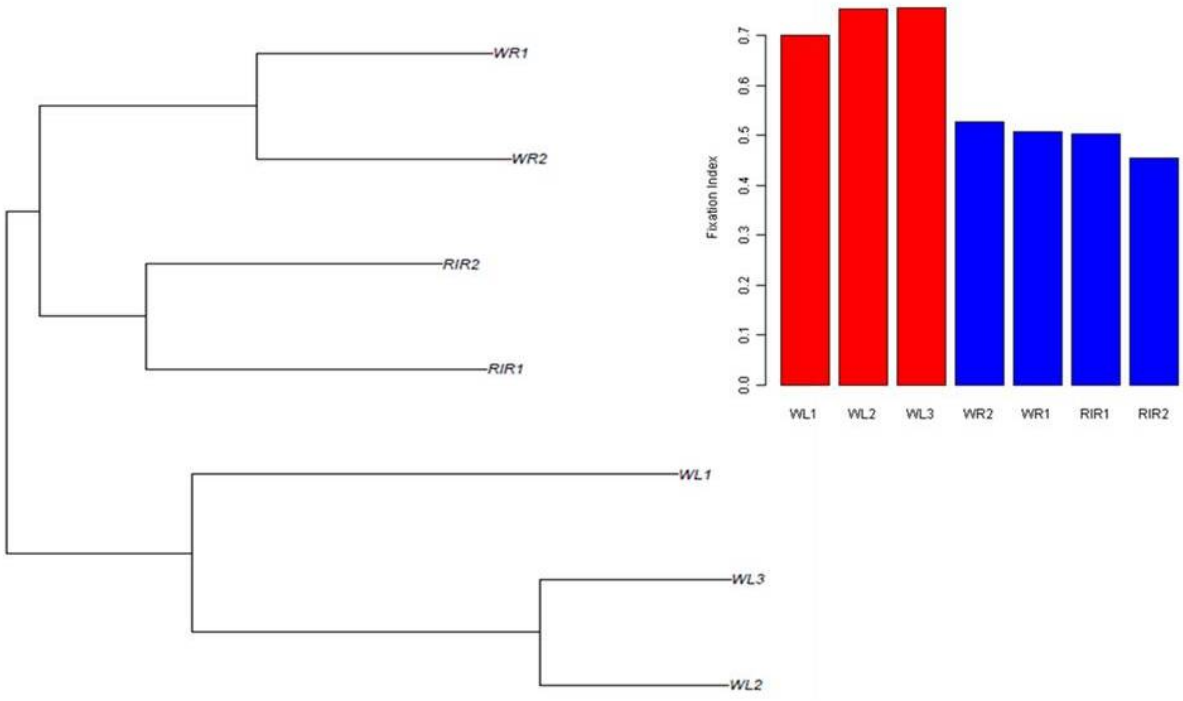

Figure 2. Reynolds' genetic distances population tree of seven commercial breeds and histogram of fixation index for each line.

FLK

Based on the FLK values distribution, a total of 107 regions (63 in all layers, 27 in white layers and 17 in brown layers) were detected as signatures of selection. All these regions were in the upper 0.05% of the distribution which is representative of regions with fixed difference between populations. The genome-wide distribution of FLK values obtained from each group - all, white and brown - are depicted in Figures 3a, 3b and 3c, respectively. Annotation was carried out for all regions with extreme FLK values, i.e. potential selection signatures. The lists of genes in selective sweeps detected with FLK are available in the supplementary tables (Table S1, S2 and S3). The annotation list is enriched with genes of biological interest involved in various pathways such as ATP metabolic process (P=0.023), metal ion binding (P=0.001), nucleic acid binding (P=0.008) and metabolic pathways (P<0.001), all of which can be related to production traits under selection in layers. The lists of pathways and gene ontologies under selection are available in supplementary tables (Table S4, S5 and S6). We identified three candidate genes which can be related to the breeding goals of chickens. *H3F3C* and *AGRP* which are associated with body growth and body weight [34,35], and *IL19* which is associated to the immune system in chicken [36]. More details about gene locations

and study groups are available in Table 2. We also detected several QTL overlapping selection signatures for traits such as breast muscle weight, abdominal fat weight and liver weight, which all are related to the breeding goals of chickens (Table 3).

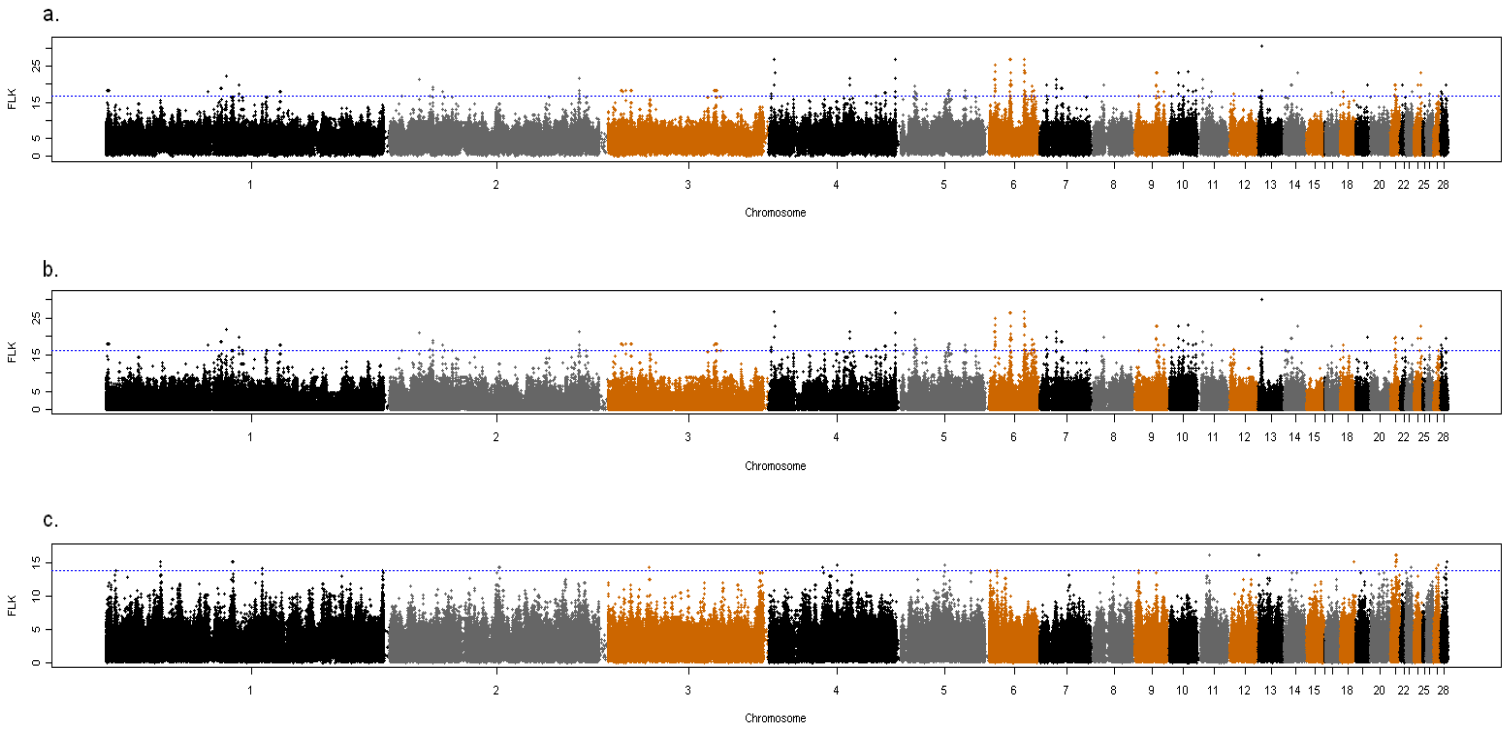

Figure 3. Manhattan plot of FLK analysis over the entire genome. Blue line indicates the upper 0.05% of FLK distribution, for (a) within all breeds, (b) within white breeds, and, (c) within brown breeds.

Table 2. Genes associated with productive traits in FLK and hapFLK analysis in all three studies. 'All', 'White', and 'Brown' stand for inclusion of all the commercial breeds, analysis within white layers and analysis within brown layers, respectively. 's' stands for similarity and 'd' for difference.

Chr	Gene	Function	Test	Group
1	SOX10	Causal mutation underlying the dark brown mutational phenotype in chickens.	hapFLK	All(d) and Brown(d)
3	H3F3C	Potential role in early feed stress responses and adaptation to feed intake stress.	FLK	All(d) , White(d)
10	IGF-1R	Associated with chicken early growth and carcass traits.	hapFLK	Brown(s)
11	AGRP	Associated with chest width, body weight, and high slaughter rate.	FLK	All(d) , White(d)
20	BPIFB8	A molecular actor of the avian egg natural defense.	hapFLK	Brown(s)
26	IL19	Associated with immunoprotection.	FLK	All(d) , White(d)
27	STAT5B	A potential genetic marker for growth and reproduction traits.	hapFLK	Brown(s)

Table 3. QTL associated with productive traits in FLK analysis in all three studies. 'All' stands for inclusion of all commercial breeds, and 'White' for analysis within white layers.

Chr	QTL	Group
1	Fear-tonic immobility duration	All, White
4	Disease-related traits	All, White
5	Disease-related traits	All, White
6	Liver weight	All, White
11	Breast muscle weight	All, White
26	Abdominal fat weight	All, White
26	Abdominal fat percentage	All, White

hapFLK

Based on the hapFLK values distribution, a total of 41 regions (17 in all layers, 12 in white layers and 12 in brown layers) were detected as selection signatures. All these regions were in either the upper or the lower 0.05% of the distribution, which represent regions with a fixed difference or fixed similarity between populations, respectively. The genome-wide distribution of hapFLK values with 5 haplotype clusters obtained for each group - all, white and brown - are depicted in Figures 4a, 4b and 4c, respectively. Annotation was carried out for all regions with extreme hapFLK values, i.e. potential selective sweeps. The lists of genes for selective sweeps detected with hapFLK are available in the supplementary tables (Table S7, S8 and S9). The annotation list is enriched with genes of biological interest involved in various pathways such as nerve development (p=0.027), growth factor receptor (p=0.008), RNA metabolic process (p=0.042) and skeletal muscle cell differentiation (p=0.032), all of which could be related to production traits indirectly. The lists of pathways and gene ontologies under which were detected under selection in this study are available in the supplementary tables (Table S10, S11 and S12). We identified four genes that were related to the breeding goals of chickens with the hapFLK method. *IGF-1R* and *STAT5B* are associated with growth and carcass traits [37,38]. *BPIFB8* and *SOX10,* which are associated with egg natural defense [39] and dark brown mutational phenotype [40] respectively (more details is available in Table 2). Several QTL, which were related to the breeding goals of egg-layer chickens were detected as well, for traits such as drumstick and thigh morphology, carcass weight and shank length. A complete list of all QTL with more details is available in Table 4.

Table 4. QTL associated with productive traits in hapFLK analysis in all three studies. 'All', 'White', and 'Brown', stand for inclusion of all the commercial breeds, analysis within white layers and analysis within brown layers, respectively. 's' stands for similarity and 'd' for difference.

Chr	QTL	Group
1	Abdominal fat percentage	All(d), Brown(d) and White(d)
1	Heart weight	All(s) and Brown(s)
2	Carcass weight	All(s) and Brown(s)
2	Drumstick and thigh weight	All(s) and Brown(s)
2	Drumstick and thigh muscle weight	All(s) and Brown(s)
2	Shank length	All(s) and Brown(s)
2	Shank circumference	All(s) and Brown(s)
2	Heart weight	White(s)
9	Liver percentage	White(s)

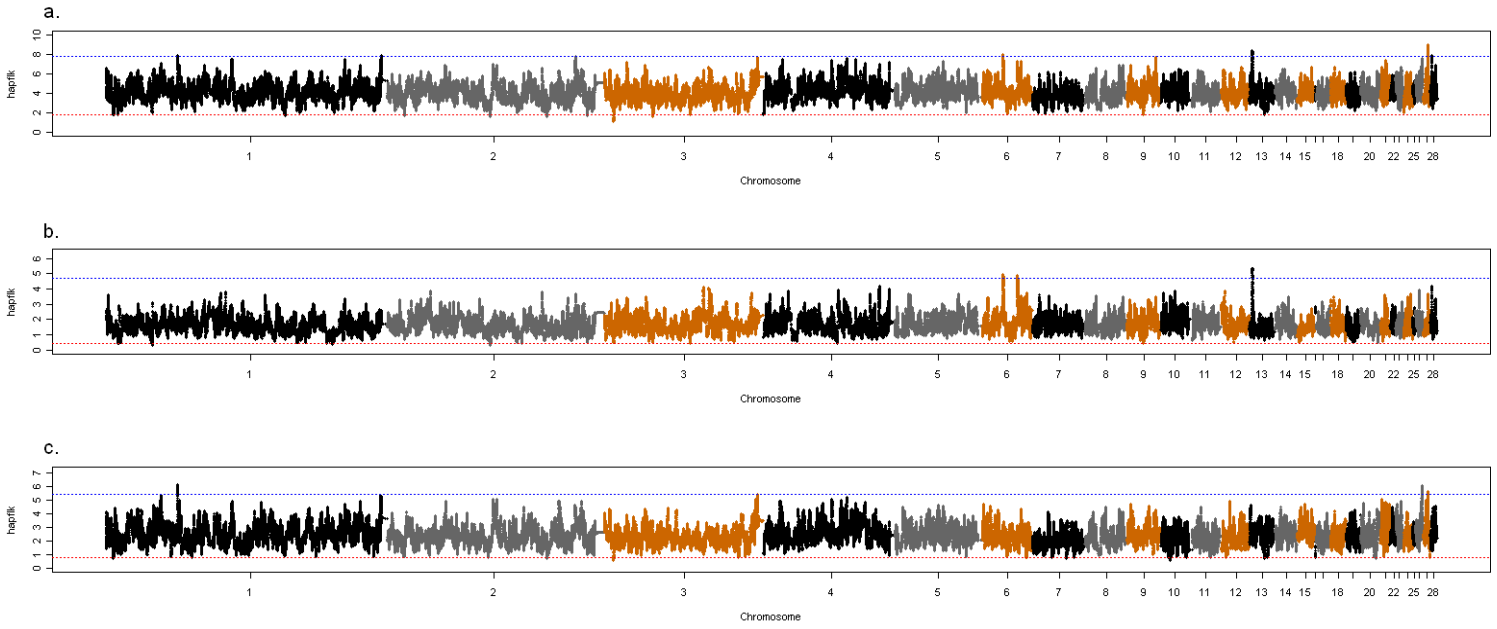

Figure 4. Manhattan plot of hapFLK analysis over the entire genome with 5 clusters. Blue (red) line indicates the upper (lower) 0.05% of hapFLK distribution, for, (a) within all breeds, (b) within white breeds, and, (c) Within brown breeds.

Discussion

Structure analysis and P_0 comparison

Our population structure analyses are largely in agreement with the expected historical origin of the breeds [21] as well as with the previous study using the same data [20].

One of the issues in the FLK and hapFLK analysis in this study is using only 4 ancestral chickens for development of the population's kinship matrix. We assessed whether using a different set of outgroup individuals could possibly change our findings by verifying the influence of the outgroup set on the estimation of the ancestral allele frequency (p_0). p_0 can be seen as a nuisance parameter in the model that has to be estimated from the data through the kinship matrix. We studied the possible impact of the number of ancestral chickens used by comparing p_0 when being calculated from 4 ancestral chickens (our ANC group) vs. 40 ancestral chickens (consisted of 20 *Gallus gallus gallus* and 20 *Gallus gallus spadiceus* which were genotyped with Axiom® Genome-Wide Chicken Genotyping Array of Affymetrix and were available only for this comparison). p_0 was calculated for each group (ANC group and 40 ancestral chickens) for every SNP on the 600K SNP chip. Pairwise comparison of each group's p_0 values along the genome gave an average correlation of 0.95. This high correlation suggests that there is no vital difference in development of population's kinship matrix with 4 or 40 ancestral chickens. Therefore the kinship matrix calculated based on four ancestral chickens, which had been genotyped for the complete set of close to one million SNPs was considered sufficient. A histogram of the differences in p_0 estimated with the two outgroup sets is shown in figure S1, showing that more than 90% of the differences are less than ± 0.02.

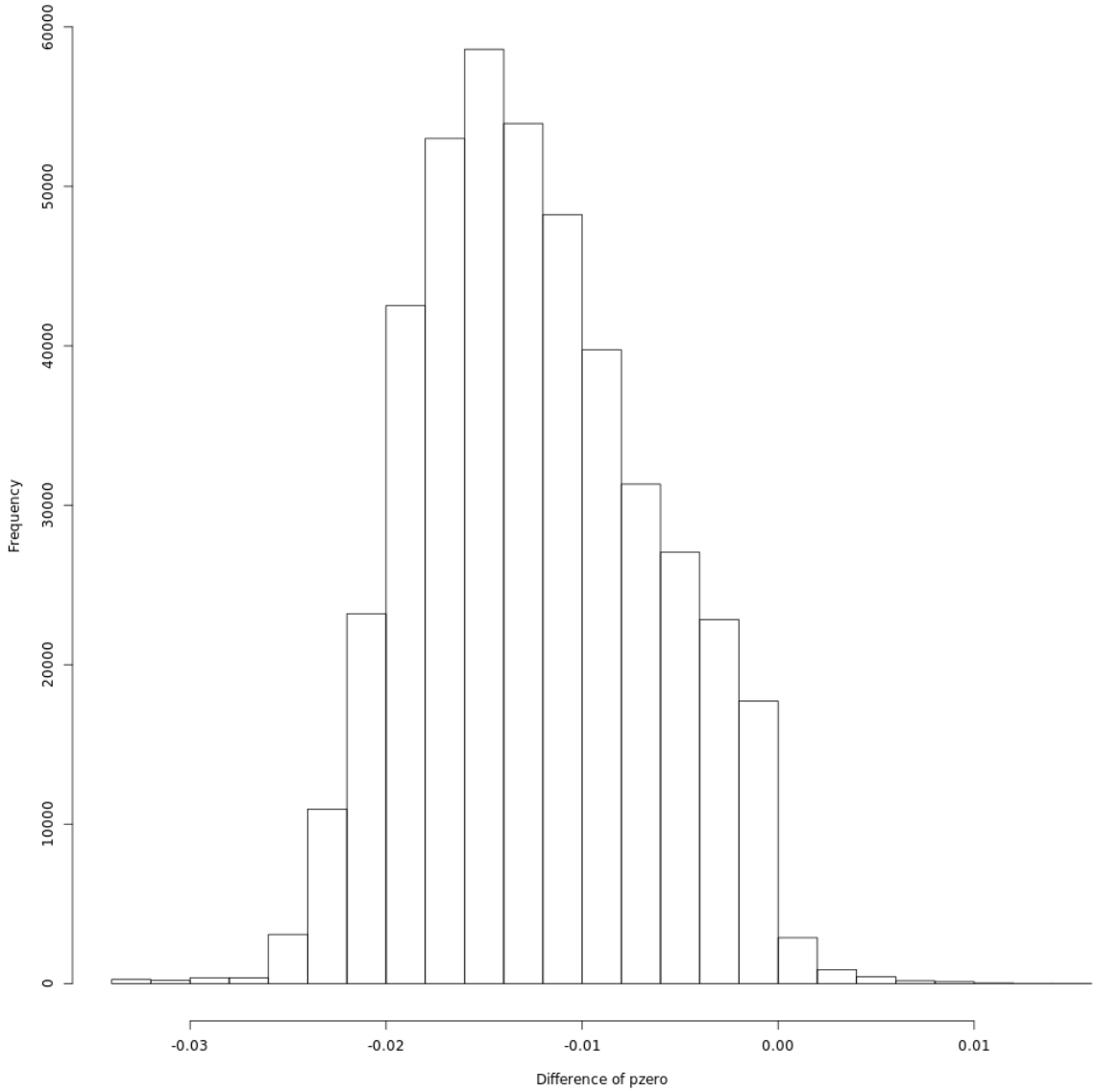

Figure 1S. Histogram of p_0 difference between the calculation with 4 ancestral chickens and 40 ancestral chickens.

Fitting of gamma distribution

Although the outlier approach is an effective and widely used method for identification of genes under selection lacking known phenotypes [41], an outlier signal is not necessarily synonymous with regions being under selection [42]. Therefore we fitted a gamma distribution to the hapFLK in order to estimate the false discovery rate (FDR). This approach suggested an FDR of 10-20% in our analysis. This is probably due to lack of power in our analysis, which, as mentioned before, could be overcome by using a larger number of populations.

F_{ST}, FLK and hapFLK

A large overlap exists between the regions that have been determined as regions under selection in a previous study with F_{ST} [20] and the current analysis of FLK and hapFLK as shown by the Venn diagram for the number of SNPs identified as being under selection with either of the methods shown in figure 5. We detected a much lower number of selection signatures with FLK (7.9%) and hapFLK (2.4%) compared to the F_{ST} based results reported in our earlier study on the same data [20]. A finding suggested that ten-thousands of polymorphisms respond to selection, which was the case in our earlier work [20], does not appear realistic [43]. Many of the outliers detected with F_{ST} must be considered as false positives, which might be partly due to the fact that the method assumes populations to have the same effective size and to have emerged independently from the same ancestral population. However, since FLK and hapFLK take different effective population sizes and hierarchical phylogenies into account, a much lower number of selection signatures were detected with these methods.

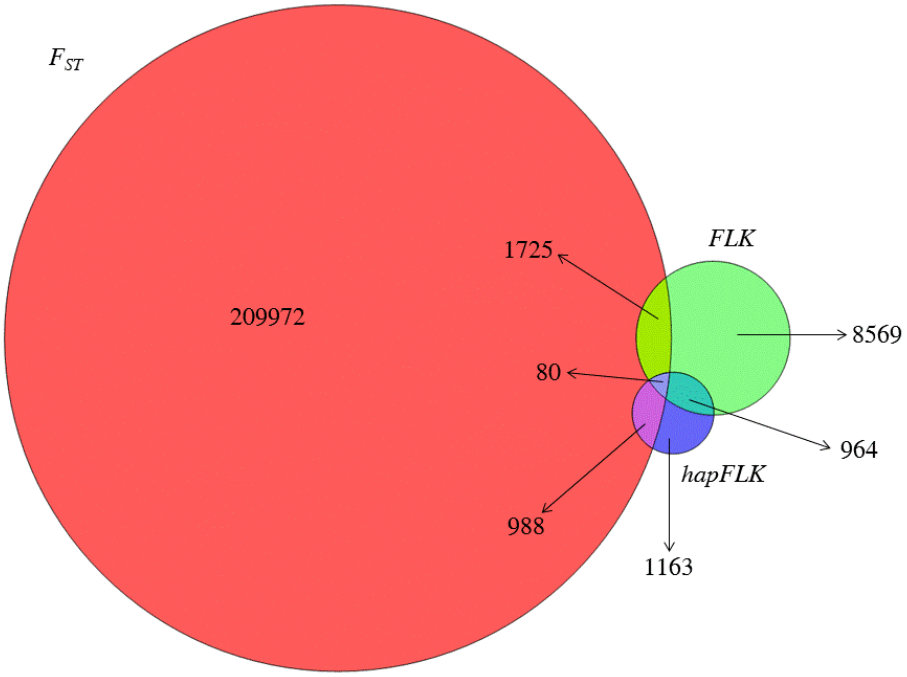

Figure 5. Venn diagram of overlapping SNPs identified as under selection, with F_{ST}, FLK and hapFLK methods.

As an example, in figure 6a we demonstrate allele frequencies at SNP positions around the *TGFB2* gene (Chr3: 18,690,003-18,753,123) which was detected as a gene under selection by F_{ST} [20] due to a reduction of diversity within the WL breed. However, since this reduction

exists only within the WL breed this can also be explained by drift alone. By taking the population tree into consideration, FLK does not detect any signals in this region. Another example is the region around the *H3F3C* gene (Chr3: 16,483,162-16,487,393) which was detected to be under selection by FLK. Allele frequencies around this region shows that a huge diversity exists between some breeds (figure 6b). We detect an outlier with FLK in particular because WR1 and WR2 show very different patterns of allele frequencies in this region although they are closely related in the population tree. However F_{ST} was not able to detect any signal here, since F_{ST} treats each population as an independent evidence for sweep detection and does not consider the huge difference between WL, RIR and WR breeds. There are as well cases in which all three methods (F_{ST}, FLK and hapFLK) were able to detect the region under selection. An example is a 60Kb region on chromosome 10 (6,799,776-6,738,610). Figure 6c shows allele frequencies around this region.

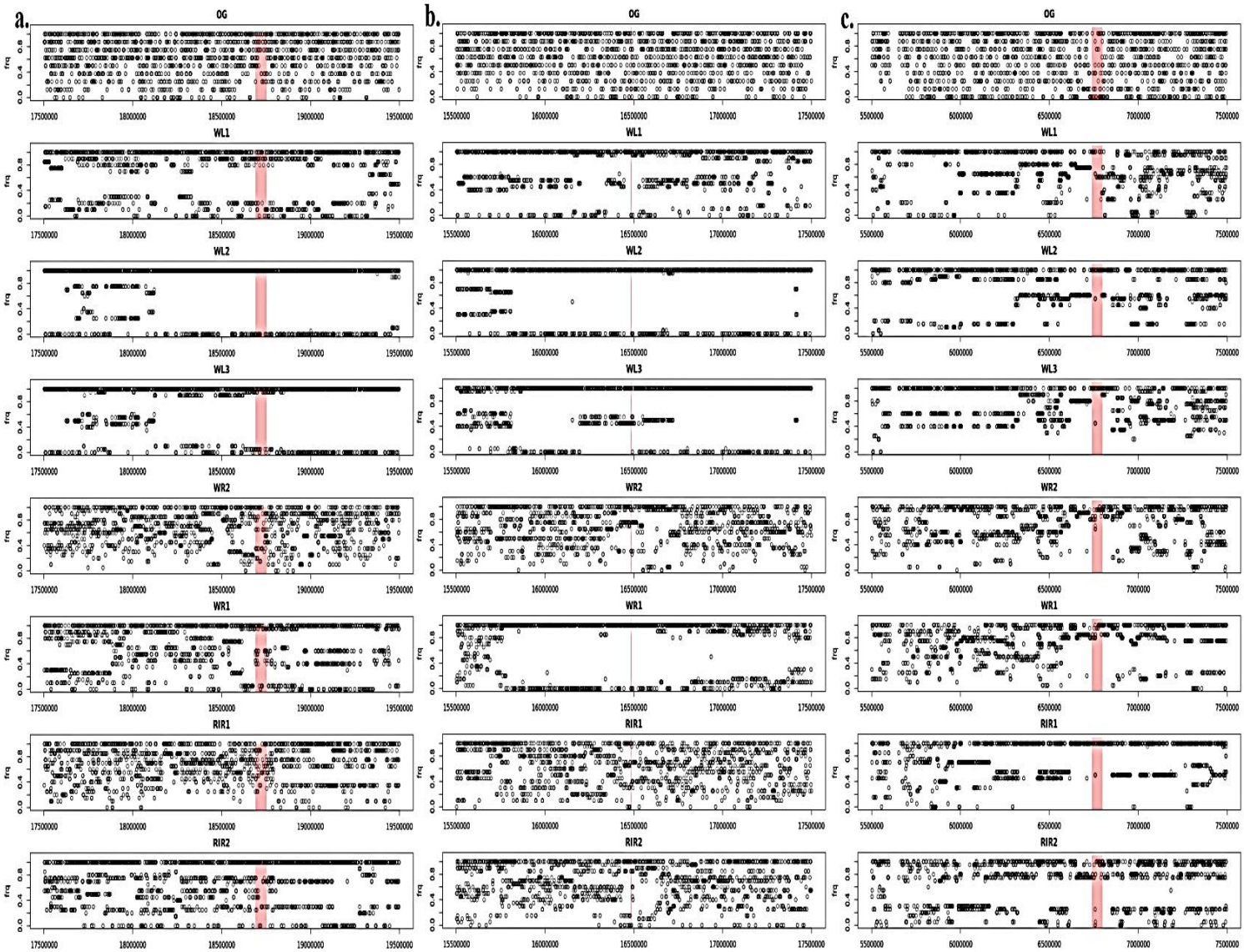

Figure 6. Allele frequency in different breeds for 2 Mbp around the intended region. Red box indicates, for (a) *TGFB2* gene (Chr3: 18,690,003-18,753,123), (b) *H3F3C* gene (Chr3: 16,483,162-16,487,393) and (c) 60Kb region on chromosome 10 (6,799,776-6,738,610).

A complete hard sweep is expected to be large [44], while a soft sweep is more likely to have smaller size [45]. In the current study we detected smaller sweeps (bp length) compared to our F_{ST} study, which may be due to the fact that hapFLK has greater power in detection of soft sweeps. Nevertheless we should as well take into account the false positive rate of our F_{ST} study. A boxplot of sweep size with F_{ST} and hapFLK method is shown in figure S2.

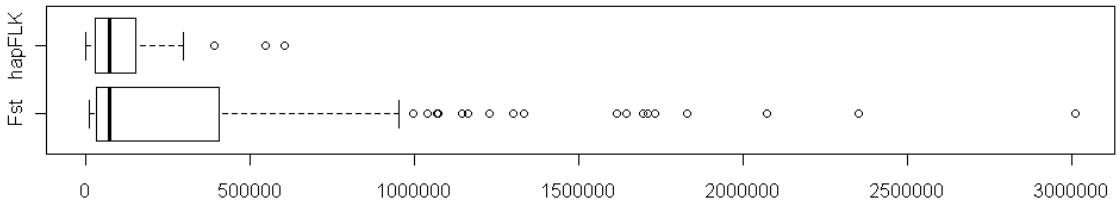

Figure 2S. Boxplot of sweep size with F_{ST} and hapFLK method.

A vast majority of differentiated polymorphisms in our data set could be caused by genetic drift. Genetic drift is high when the (effective) population size is small [46] which is the case in commercial laying breeds [47]. Since regions differentiated by selection and regions differentiated by drift alone may overlap, there is a lack of power in our analysis. This could be solved by using a larger number of populations to minimize the risk that a systematic pattern of differentiation in many breeds (say, several white layers vs. several brown layers) is created at random by drift alone. However, we detected several genes related to the breeding goals of egg-layer chickens, such as low body weight, high reproduction performance and good feed conversion [48], both with FLK and hapFLK. For instance, with the FLK method we detected several QTL associated to disease-related traits and breast muscle weight, as well as *AGRP* (agouti related protein homolog), which is associated with breast muscle water loss rate, chest width, body weight, slaughter rate and semi-evisceration weight [35].

In the hapFLK analysis, we also detected several genes, which are associated with growth and carcass traits, such as *IGF-1R* and *STAT5B. STAT5B* (signal transducer and activator of transcription 5B) is associated with growth and reproduction traits [38]. *IGF-1R* (insulin-like growth factor 1) is similar to *IGF2* [49], which was detected in our previous work [20]. *IGF-1R* is associated with chicken early growth and carcass traits [37]. We additionally detected several QTL associated to carcass weight, drumstick weight and shank length. QTL associated with meat production, as well as both *IGF-1R* and *STAT5B,* were located in regions that were similar between brown layers. Supporting results were found in our previous study [20], where we detected genes associated to meat quality and production in brown layers, which reflects the fact that brown egg-layers were originally a dual-purpose breed [21].

Bonhomme *et al.* (2010) [12] and Fariello *et al.* (2013) [15] showed with simulation that using FLK or hapFLK method to detect selection signatures in comparison to other F_{ST}-like approaches greatly increases the detection power. Specifically, hapFLK statistic has more power in detecting sweeps occurring in several populations. Due to this, we were able to detect *SOX10* with hapFLK which was not detected by F_{ST} or FLK method. *SOX10* is a gene on chromosome one underlying the dark brown mutational phenotype in chickens plumage [40]. *SOX10* was detected in regions that were different between brown layers. Re-estimation of the local tree using haplotype clusters frequencies (figure 7a) and haplotype frequencies (figure 7b) for the region surrounding *SOX10* revealed selection in the RIR breeds in this region. RIR is the only breed with dark brown feather in our data set [50], which is in great agreement with our selection signature detection.

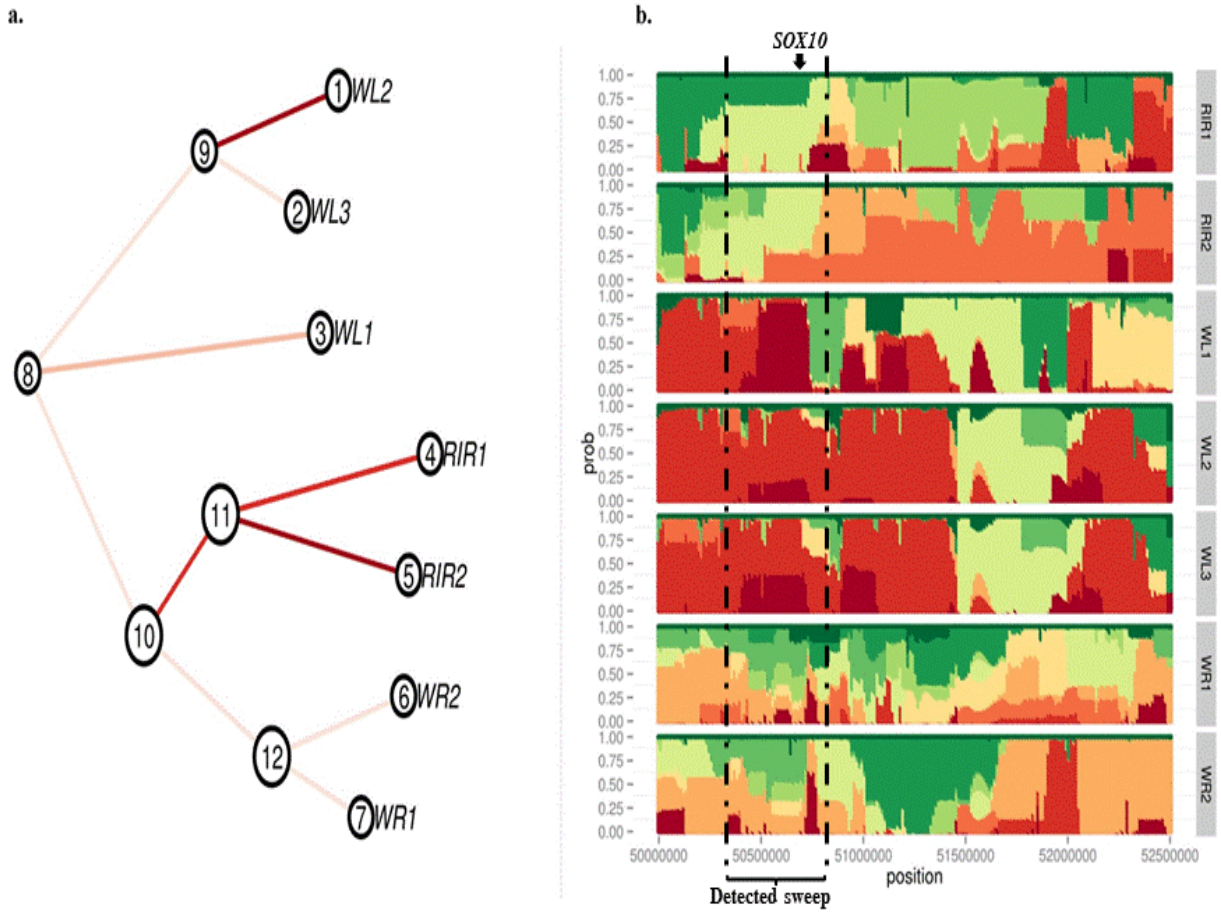

Figure 7. (a) Re-estimation of local tree using haplotype clusters frequencies for surrounding region of *SOX10* gene. (b) Haplotype frequencies for the surrounding region of *SOX10* gene (50.8 Mbp).

Conclusions

In conclusion we were able to identify several putative selection signature regions with genes corresponding to the productions traits. Some of these annotated genes were similar (or had similar functions) to our findings in our previous work [20]. However, several of the detected regions were not associated with any genes related to production traits, which could be due to insufficient knowledge about these regions [51]. We did not identify selection signatures that were reported in other studies on chicken [17,52] which could be due to lack of diversity in our data compared to their data set. By detection of *SOX10* as a gene under selection, we demonstrated that the use of haplotype frequencies and consideration of hierarchical structure can improve the power of detection in our data set.

Acknowledgement

This research was funded by the German Federal Ministry of Education and Research within the AgroClustEr "Synbreed – Synergistic plant and animal breeding" (Funding ID: 0315528C). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Dr. Preisinger is the head of Genetic department of LOHMANN Tierzucht GMBH. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials. We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Göttingen University.

Supporting Information

Supplementary Table 1. List of genes for selective sweeps detected with FLK with 0.05% threshold in all layers.

Chr	Start	End	Description	FLK
1	621303	630454	interferon regulatory factor 5	18.18
1	630567	647745	transportin 3	18.18
1	658443	664773	Smoothened homolog	18.18
1	690901	708802	Adenosylhomocysteinase	18.18
1	710117	727850	striatin interacting protein 2	18.18
1	757267	795092	nuclear respiratory factor 1	18.18
1	807182	849202	ubiquitin-conjugating enzyme E2 H	18.18
1	850460	859444	nuclear-interacting partner of ALK	18.18
1	911097	915397	carboxypeptidase A5 precursor	18.18
1	916281	920045	carboxypeptidase A1 preproprotein	18.18
1	921538	929838	Centrosomal protein of 41 kDa	18.18
1	935413	954122	Coatomer subunit gamma	18.18
1	995787	1011804	ankyrin repeat domain-containing protein 16	18.18
1	1012026	1028924	rab GDP dissociation inhibitor beta	18.18
1	1016281	1061797	family with sequence similarity 208, member B	18.18
1	1068473	1076747	ankyrin repeat and SOCS box protein 13	18.18
1	1088060	1099472	neuroepithelial cell-transforming gene 1 protein	18.18
1	1361642	1822074	exocyst complex component 4	18.03
1	71808627	71817151	BCL2-like 14 (apoptosis facilitator)	18.03
1	71833436	71943987	low density lipoprotein receptor-related protein 6	18.03
1	71974273	71980456	MANSC domain containing 1 precursor	18.03
1	71983303	72042978	Loss of heterozygosity 12 chromosomal region 1 protein homolog	18.03
1	72046586	72113670	dual specificity phosphatase 16	18.03
1	72156810	72166175	G protein-coupled receptor 19	18.03
1	81373015	81656000	Limbic system-associated membrane protein	18.99
1	81680485	81734254	growth associated protein 43	18.99
1	94474405	94635350	glucan (1,4-alpha-), branching enzyme 1	18.03
1	122962850	123267815	FERM and PDZ domain containing 4	18.03
1	123394258	123418544	male-specific lethal 3 homolog (Drosophila)	18.03
2	20547214	20632812	multidrug resistance protein 1	21.33
2	20656627	20694939	RUN domain-containing protein 3B	21.33
2	20696196	20709565	solute carrier family 25, member 40	21.33
2	20711043	20726377	DBF4 homolog (S. cerevisiae)	21.33
2	20865806	20874342	sorcin	21.33
2	30537280	30564128	Sp4 transcription factor	19.22
2	30690282	30698829	cell division cycle-associated 7-like protein	19.22
2	30757166	30909032	Rap guanine nucleotide exchange factor (GEF) 5	19.22
2	30683539	30683661	U5 spliceosomal RNA	19.22
2	37711497	37756128	ubiquitin-conjugating enzyme E2E 1	18.03
2	37769194	37773772	Ribosomal protein L15	18.03
2	37787391	37809859	nuclear receptor subfamily 1 group D member 2	18.03
2	37844468	37965174	Thyroid hormone receptor beta	18.03

2	37698436	37698610	U1 spliceosomal RNA	18.03
2	134048833	134252747	trichorhinophalangeal syndrome I	21.66
2	134628301	134709262	Eukaryotic translation initiation factor 3 subunit H	21.66
3	9262357	9405997	WD repeat containing planar cell polarity effector	18.18
3	9406399	9417141	Malate dehydrogenase, cytoplasmic	18.18
3	9438436	9460691	UTP--glucose-1-phosphate uridylyltransferase	18.18
3	9462114	9489472	vacuolar protein sorting 54 homolog (S. cerevisiae)	18.18
3	9521292	9539167	protein pellino homolog 1	17.99
3	10359052	10449924	Meis homeobox 1	17.99
3	12844082	12878172	Delta-like protein	18.18
3	12887313	12960238	UPF0492 protein C20orf94 homolog	18.18
3	12961577	12970854	McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin	18.18
3	12995964	13027937	Synaptosomal-associated protein 25	18.18
3	16172663	16238767	chromogranin B (secretogranin 1)	18.18
3	16255281	16265499	DNA helicase MCM8	18.18
3	16271778	16276722	cardiolipin synthase 1	18.18
3	16360689	16391192	poly	18.18
3	16404585	16432600	protein lin-9 homolog	18.18
3	16453958	16471806	Golgi resident protein GCP60	18.18
3	16482887	16487420	histone H3.2	18.18
3	16515941	16518570	left-right determination factor 2 precursor	18.18
3	16598307	16603854	signal recognition particle 9 kDa protein	18.22
3	74490287	74540828	mitogen-activated protein kinase kinase kinase 7	18.22
3	74768582	74811042	BTB and CNC homology 1, basic leucine zipper transcription factor 2	18.22
3	74853141	74946896	Midasin	18.22
3	74953170	74985332	ankyrin repeat domain 6	18.22
3	74823486	74824862	gap junction protein, alpha 10, 62kDa	18.22
3	76289510	76312354	synaptotagmin binding, cytoplasmic RNA interacting protein	18.35
3	76326433	76373661	sorting nexin-14	18.35
3	76271065	76271131	Small nucleolar RNA SNORD50	18.35
3	76273207	76273277	Small nucleolar RNA SNORD50	18.35
4	1703060	1742346	Midline 2; Uncharacterized protein	17.37
4	1770827	1772689	TSC22 domain family, member 3	17.37
4	1790749	1795411	fatty acid amide hydrolase 2	17.37
4	1801657	1806508	chloride intracellular channel protein 2	17.37
4	1813109	1815939	G2/mitotic-specific cyclin-B3	17.37
4	1830442	1832824	shroom family member 4	17.37
4	1838457	1841000	bone morphogenetic protein 15 precursor	17.37
4	1858232	1866719	Histone deacetylase	17.37
4	1867724	1871061	40S ribosomal protein S4	17.37
4	1871906	1874674	RNA binding motif protein 41	17.37
4	1927710	1944396	FERM and PDZ domain containing 3	17.37
4	1950342	1960665	Ribose-phosphate pyrophosphokinase	17.37
4	1977381	1981476	myelin proteolipid protein	17.37
4	2006952	2017876	tyrosine-protein kinase BTK	17.37
4	2030171	2032347	aryl hydrocarbon receptor interacting protein-like 1	17.37
4	2035986	2044107	dystrophin related protein 2	17.37
4	2058474	2072715	Centromere protein I	17.37
4	2094842	2104015	H/ACA ribonucleoprotein complex subunit 4	17.37

4	2104399	2118953	55 kDa erythrocyte membrane protein	17.37
4	2122995	2142344	coagulation factor VIII, procoagulant component	17.37
4	2148375	2154417	FUN14 domain containing 2	17.37
4	2156282	2160713	C-x(9)-C motif containing 4 homolog (S. cerevisiae)	17.37
4	2163986	2168244	BRCA1/BRCA2-containing complex, subunit 3	17.37
4	2177525	2195453	Putative uncharacterized protein	17.37
4	2100634	2100709	Small nucleolar RNA SNORD83	17.37
4	2098331	2098464	Small nucleolar RNA SNORA36 family	17.37
4	2097884	2097959	Small nucleolar RNA SNORD83	17.37
4	2099885	2100131	Small nucleolar RNA SNORD17	17.37
4	2101395	2101544	Small nucleolar RNA SNORA56	17.37
4	2097097	2097343	Small nucleolar RNA SNORD17	17.37
4	4161930	4176418	solute carrier family 9, subfamily A, member 6	27.01
4	4206227	4209911	four and a half LIM domains 1	27.01
4	4214210	4248185	MAP7 domain containing 3	27.01
4	4280658	4292346	G protein-coupled receptor 112	27.01
4	4301225	4305523	bombesin receptor subtype-3	27.01
4	4345180	4349228	CD40 ligand CD40 ligand, membrane form CD40 ligand, soluble form	27.01
4	4354214	4383568	Rho guanine nucleotide exchange factor 6	27.01
4	4390057	4400409	heterogeneous nuclear ribonucleoprotein G	27.01
4	4398822	4398886	Small nucleolar RNA SNORD61	27.01
4	57404743	57409306	pituitary homeobox 2	21.67
4	57434843	57469617	glutamyl aminopeptidase (aminopeptidase A)	21.67
4	57576632	57597173	Elongation of very long chain fatty acids protein 6	21.67
4	57631764	57683454	pro-epidermal growth factor precursor	21.67
4	57708911	57722892	visual pigment-like receptor peropsin	21.67
4	57731319	57744001	complement factor I	21.67
4	57745870	57751937	phospholipase A2, group XIIA	21.67
4	57754986	57763766	caspase-6	21.67
4	57807176	57923545	PDZ and LIM domain protein 5	17.74
4	82408395	82444435	max dimerization protein 4	17.74
4	82459166	82544586	polymerase (DNA directed) nu	17.74
4	82738354	82761647	negative elongation factor A	17.74
4	82783104	82842064	Wolf-Hirschhorn syndrome candidate 1	26.79
4	89402351	89584157	exocyst complex component 6B	26.79
5	9538613	9582720	LIM domain only 1 (rhombotin 1)	19.60
5	9614317	9626415	resistance to inhibitors of cholinesterase 3 homolog	19.60
5	9634001	9761468	tubby homolog (mouse)	19.60
5	9768005	9802730	ras-related protein R-Ras2	19.60
5	9834069	9848544	Coatomer subunit beta	19.60
5	9851015	9858153	proteasome subunit alpha type-1	19.60
5	9876942	9951911	cGMP-inhibited 3,5-cyclic phosphodiesterase B	17.37
5	10549851	10784968	SRY (sex determining region Y)-box 6	17.37
5	33101816	33216460	serine/threonine-protein kinase D1	18.35
5	33413582	33433671	G2/M phase-specific E3 ubiquitin-protein ligase	18.35
5	33438355	33483021	sec1 family domain-containing protein 1	18.35
5	33517885	33534657	Cochlin	18.35
5	33543459	33598778	striatin, calmodulin binding protein 3	18.18
5	34050021	34304671	A kinase (PRKA) anchor protein 6	18.18

5	34375148	34959900	neuronal PAS domain protein 3	18.22
5	44981707	44983006	homeobox protein goosecoid	18.22
5	45092963	45128794	endoribonuclease Dicer	18.22
5	45165017	45184249	calmin (calponin-like, transmembrane)	18.22
5	45270524	45300174	spectrin repeat containing, nuclear envelope family member 3	18.22
5	45330738	45337464	glutaredoxin-related protein 5, mitochondrial	18.22
5	45328658	45328930	Small Cajal body specific RNA 13	18.22
6	3953643	3956579	lung lectin precursor	25.21
6	3961091	3967496	surfactant, pulmonary-associated protein A1 precursor	25.21
6	3981178	3989358	soluble mannose-binding lectin precursor	25.21
6	4283164	4363396	ret proto-oncogene precursor	25.21
6	14667844	14924979	adenosine kinase	26.81
6	14941845	14956153	AP-3 complex subunit mu-1	26.81
6	14961440	14990291	Vinculin	26.81
6	15060198	15068452	urokinase-type plasminogen activator preproprotein	26.81
6	16111040	16115551	zona pellucida sperm-binding protein 4	19.84
6	16273210	16314303	Phosphoinositide 3-kinase adapter protein 1	19.84
6	16318282	16329867	nucleolar and coiled-body phosphoprotein 1	19.84
6	16349917	16354551	ELOVL fatty acid elongase 3	19.84
6	16379333	16476177	golgi brefeldin A resistant guanine nucleotide exchange factor 1	19.84
6	16488268	16493350	nuclear factor NF-kappa-B p100 subunit	19.60
6	24664740	24940362	sortilin-related VPS10 domain containing receptor 1	19.60
6	25264735	25352280	gamma-adducin	19.60
6	30060586	30109050	phosphatidic acid phosphatase type 2 domain containing 1A	19.60
6	30185811	30223053	WD repeat domain 11	19.60
6	30411293	30490016	Fibroblast growth factor receptor 2	19.60
6	30550113	30621520	arginyl-tRNA--protein transferase 1	18.35
6	31636707	31661235	family with sequence similarity 53, member B	18.35
6	31686516	31702105	BRISC complex subunit Abro1	18.35
6	32007895	32012786	testis expressed 36	18.35
6	32058339	32075635	DEAH (Asp-Glu-Ala-His) box polypeptide 32	18.35
7	4312646	4330369	bifunctional purine biosynthesis protein PURH	19.92
7	4333468	4382817	fibronectin precursor	19.92
7	4527554	4590262	NEDD8-conjugating enzyme UBE2F	19.92
7	4591385	4595864	receptor (G protein-coupled) activity modifying protein 1	19.92
7	4611641	4662295	leucine rich repeat (in FLII) interacting protein 1	19.92
7	4722132	4740447	melanophilin	19.92
7	4752831	4805577	collagen alpha-3(VI) chain precursor	19.92
7	11536104	11616253	pleckstrin homology domain containing, family M, member 3	21.49
7	11686741	11704137	cyclic AMP-responsive element-binding protein 1	21.49
7	11766883	11819797	Kruppel-like factor 7 (ubiquitous)	21.49
7	11854607	11868058	carboxypeptidase O	21.49
7	11950838	12015748	disintegrin and metalloproteinase domain-containing protein 23	21.49
7	12072509	12076665	elongation factor 1-beta	21.49
7	12077169	12090936	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	21.49
7	12107101	12129187	INO80 complex subunit D	21.49
7	12074025	12074158	Small nucleolar RNA SNORA41	21.49
7	12074705	12074740	Small nucleolar RNA Z196/R39/R59 family	21.49
7	15219971	15271809	alkylglycerone phosphate synthase	18.74

7	15304546	15320356	nuclear factor erythroid 2-related factor 2	18.74
8	7564281	7592106	smg-7 homolog, nonsense mediated mRNA decay factor (C. elegans)	19.92
8	7604691	7609670	actin-related protein 2/3 complex subunit 5	19.92
8	7641619	7708062	ral guanine nucleotide dissociation stimulator-like 1	19.92
8	7903724	7931488	ER degradation enhancer, mannosidase alpha-like 3	19.92
8	7933908	7992849	protein Niban	19.92
8	7564912	7565074	U1 spliceosomal RNA	19.92
9	15087461	15106244	eukaryotic translation initiation factor 4E family member 2	23.24
9	15246615	15264143	eukaryotic translation initiation factor 2B, subunit 5 epsilon, 82kDa	23.24
9	15281328	15292012	dishevelled, dsh homolog 3 (Drosophila)	23.24
9	15293532	15316234	AP-2 complex subunit mu	23.24
9	15332834	15340129	von Willebrand factor A domain containing 5B2	23.24
9	15340601	15343371	ALG3, alpha-1,3- mannosyltransferase	23.24
9	15356178	15362770	endothelin converting enzyme 2	23.24
9	15363725	15369726	26S proteasome non-ATPase regulatory subunit 2	23.24
9	15371232	15387501	eukaryotic translation initiation factor 4 gamma, 1	23.24
9	15395572	15405268	chloride channel, voltage-sensitive 2	23.24
9	15406844	15413438	chordin precursor	23.24
9	15419502	15421617	thrombopoietin precursor	23.24
9	15437757	15449465	EPH receptor B3	23.24
9	15380309	15380385	Small nucleolar RNA SNORD66	23.24
9	15381519	15381596	Small nucleolar RNA SNORD66	23.24
9	16565596	16587667	fragile X mental retardation syndrome-related protein 1	17.88
9	16663283	16672397	tetratricopeptide repeat protein 14	17.88
9	16760382	16850020	peroxisomal biogenesis factor 5-like	17.88
9	16861502	16902461	Ubiquitin carboxyl-terminal hydrolase 13	17.88
9	16903953	16907474	NADH dehydrogenase	17.88
9	16911951	16921482	actin-like 6A	17.88
9	16984938	17003984	mitofusin-1	17.88
9	16820695	16820798	U6 spliceosomal RNA	17.88
9	20419761	20464371	cholinesterase precursor	17.99
9	20586079	20587887	SLIT and NTRK-like family, member 3	17.99
10	6434758	6619897	disintegrin and metalloproteinase domain-containing protein 10 precursor	23.24
10	6572777	6596733	aquaporin 9	23.24
10	6604818	6661741	Retinal dehydrogenase 2	23.24
10	6816944	6857906	cingulin-like 1	23.24
10	6903270	7058903	transcription factor 12	19.38
10	9545024	9553009	solute carrier family 24, member 5 precursor	19.38
10	9551752	9570906	myelin expression factor 2	19.38
10	9587065	9632792	solute carrier family 12 (sodium/potassium/chloride transporters), member 1	19.38
10	17122370	17146830	proprotein convertase subtilisin/kexin type 6	18.03
10	17199826	17220109	ankyrin repeat and death domain containing 1A	18.27
10	18440729	18553284	dual specificity mitogen-activated protein kinase kinase 5	18.27
10	18595984	18624673	E3 SUMO-protein ligase PIAS1	18.27
10	18632517	18638897	ceroid-lipofuscinosis, neuronal 6, late infantile, variant	18.27
10	18641302	18648029	protein fem-1 homolog B	18.27
10	18699925	18729841	coronin, actin binding protein, 2B	18.27
10	18770495	18779187	NADPH oxidase 5	18.27
10	18807760	18817356	glucuronic acid epimerase	18.27

10	18823404	18826891	progestin and adipoQ receptor family member V	18.27
10	18828578	18846365	kinesin-like protein KIF23	18.27
11	983068	995051	E2F transcription factor 4; Uncharacterized protein	21.49
11	995632	1005522	engulfment and cell motility 3	21.49
11	1011955	1025358	KIAA0895-like	21.49
11	1029318	1036713	lecithin-cholesterol acyltransferase	21.49
11	1073882	1097277	family with sequence similarity 65, member A	21.49
11	1102641	1128402	transcriptional repressor CTCF	21.49
11	1130092	1147360	RGD motif, leucine rich repeats	21.49
11	1165618	1177771	alanyl-tRNA synthetase, cytoplasmic	21.49
11	1189504	1194020	fibulin 7	21.49
11	1213866	1215689	tubulin polymerization-promoting protein family member 3	21.49
11	1223848	1237337	potassium channel tetramerisation domain containing 19	21.49
11	1240508	1302916	pleckstrin homology domain containing, family G member 4	21.49
11	1336785	1351404	FH1/FH2 domain-containing protein 1	21.49
11	1389699	1391495	agouti-related protein precursor	21.49
11	1398993	1402340	N-lysine methyltransferase SETD6	21.49
11	1405351	1453089	CCR4-NOT transcription complex, subunit 1	21.49
11	1424514	1424649	Small nucleolar RNA SNORA46	21.49
11	1432730	1432863	Small nucleolar RNA SNORA76	21.49
11	7596758	7603866	Cytochrome b-c1 complex subunit Rieske, mitochondrial	17.99
11	7841312	7844172	processing of precursor 4	17.99
11	7933476	7946064	G1/S-specific cyclin-E1	17.99
11	7961122	8028603	URI1, prefoldin-like chaperone	17.99
12	2286271	2288801	guanylate binding protein	17.20
12	2342130	2383187	ring finger protein 123	17.20
12	2383108	2385611	GDP-mannose pyrophosphorylase B	17.20
12	2420987	2427304	cadherin-related family member 4	17.20
12	2427451	2434824	ubiquitin-like modifier activating enzyme 7	17.20
12	2435916	2437477	aminomethyltransferase, mitochondrial precursor	17.20
12	2441415	2452604	dystroglycan precursor	17.20
12	2479299	2483435	hepatocyte growth factor-like protein precursor	17.20
12	2487790	2491678	acylamino-acid-releasing enzyme	17.20
12	2494176	2539876	bassoon presynaptic cytomatrix protein	17.20
12	2549854	2567887	TRAF-interacting protein	17.20
12	2568476	2571564	CaM kinase-like vesicle-associated	17.20
12	2579806	2587122	macrophage-stimulating protein receptor precursor	17.20
12	2588483	2592570	MON1 homolog A (yeast)	17.20
12	2598043	2648508	RNA-binding protein 6	17.20
12	2659272	2697071	RAD54-like 2 (S. cerevisiae)	17.20
12	2749644	2795258	Fanconi anemia, complementation group D2	17.20
12	2379402	2380922	adhesion molecule with Ig-like domain 3	17.20
13	2010626	2012530	leucine rich repeat transmembrane neuronal 2	30.64
13	2134785	2156387	stress-70 protein, mitochondrial precursor	30.64
13	2146299	2146368	Small nucleolar RNA SNORD63	30.64
14	5075181	5086183	DNA topoisomerase 3-alpha	19.72
14	5112713	5127320	phosphoribosyl pyrophosphate synthase-associated protein 2	19.72
14	5129153	5157900	solute carrier family 5, member 10	19.72
14	5138381	5150303	family with sequence similarity 83, member G	19.72

14	5223282	5225616	B9 protein domain 1	19.72
14	5230050	5338554	Voltage-dependent T-type calcium channel subunit alpha-1H	19.72
14	9934515	9974931	ubiquitin carboxyl-terminal hydrolase 7	23.08
14	10046044	10052775	phosphomannomutase 2	23.08
14	10103934	10114456	methyltransferase-like protein 22	23.08
17	4788861	4790491	ovoglycoprotein precursor	17.74
17	4820725	4824113	nuclear apoptosis inducing factor 1	17.74
17	4829664	4860591	early estrogen-induced gene 1 protein	17.74
17	4865441	4867919	dolichol phosphate-mannose biosynthesis regulatory protein	17.74
17	4870177	4871635	alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1	17.74
17	4872095	4874530	ST6 -N-acetylgalactosaminide alpha-2,6-sialyltransferase 6	17.74
17	4877410	4880796	Adenylate kinase isoenzyme 1	17.74
17	4882932	4891674	endoglin precursor	17.74
17	4892549	4896134	folylpolyglutamate synthase	17.74
17	4896492	4902733	Cyclin-dependent kinase 9	17.74
17	4903795	4923543	SH2 domain containing 3C	17.74
17	4929079	4931968	tetratricopeptide repeat domain 16	17.74
17	4933496	4946785	cerebral endothelial cell adhesion molecule	17.74
17	4949346	4967354	Ubiquitin-related modifier 1 homolog	17.74
17	4983785	4987536	coenzyme Q4 homolog (S. cerevisiae)	17.74
17	5016759	5075578	dynamin 1	17.74
18	2020996	2066954	Netrin-1	18.03
18	2070106	2146328	syntaxin-8	18.03
18	2180734	2183708	ubiquitin specific peptidase 43	18.03
18	2188691	2205886	Rho GTPase activating protein 44	18.03
18	2223389	2230202	uncharacterized protein LOC417324	18.03
19	7951502	7972199	protein phosphatase, Mg2+/Mn2+ dependent, 1D	19.92
19	7979200	7997166	amyloid protein-binding protein 2	19.92
19	8096015	8104574	carbonic anhydrase IV	19.92
19	8148378	8165006	Gametogenetin-binding protein 2	19.92
19	8166412	8187734	Dehydrogenase/reductase SDR family member 11	19.92
21	3252441	3261217	G protein-coupled receptor 157	19.92
21	3470173	3492933	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta isoform	19.92
21	3493075	3527740	calsyntenin-1 precursor	19.92
21	3573056	3581801	Protein LZIC	19.92
21	3581913	3588648	nicotinamide nucleotide adenylyltransferase 1	19.92
21	3590985	3592942	retinol binding protein 7, cellular	19.92
21	3595756	3630094	ubiquitination factor E4B	19.92
21	3639974	3713856	kinesin family member 1B	19.92
21	3719439	3729173	6-phosphogluconate dehydrogenase, decarboxylating	19.92
21	3705900	3706069	TUC338	19.92
22	1256816	1265197	Charged multivesicular body protein 7	19.72
22	1270094	1276295	tumor necrosis factor receptor superfamily member 10B precursor	19.72
22	1343627	1375329	exportin-7	19.72
22	1376680	1381157	docking protein 2, 56kDa	19.72
22	1397665	1418597	GDNF family receptor alpha-2 precursor	19.72
22	1597651	1602693	TELO2 interacting protein 2	19.72
23	4570631	4804042	CUB and Sushi multiple domains 2	18.03
23	4824370	4834663	collagen, type IX, alpha 2	18.03

23	4835192	4847417	small ArfGAP2	18.03
23	4901524	4907272	potassium voltage-gated channel, KQT-like subfamily, member 4	18.03
23	4910836	4916448	tubulointerstitial nephritis antigen-like 1	18.03
23	4918792	4920966	penta-EF-hand domain containing 1	18.03
23	4921631	4939423	collagen, type XVI, alpha 1	18.03
23	4943828	4960852	brain-specific angiogenesis inhibitor 2	18.03
24	2427011	2464846	immunoglobulin superfamily, member 9B	19.95
24	2516848	2542717	non-SMC condensin II complex, subunit D3	19.95
24	2553386	2558481	Thymocyte nuclear protein 1	19.95
24	2845500	2863118	potassium-transporting ATPase alpha chain 2	19.95
24	2902288	3002638	GRAM domain containing 1B	23.08
24	4392700	4566947	ubiquitin conjugation factor E4 A	23.08
24	4557911	4561306	RNA-binding protein 7	23.08
24	4664469	4732605	cell adhesion molecule 1	23.08
26	2504099	2506388	Interleukin-10	17.34
26	2517673	2519712	interleukin 19	17.34
26	2520032	2531914	polymeric immunoglobulin receptor precursor	17.34
26	2556703	2560172	ubiquitin thioesterase OTU1	17.34
26	2576519	2583374	C4b-binding protein alpha chain precursor	17.34
26	2599568	2605277	complement component (3b/4b) receptor 1-like precursor	17.34
26	2627129	2639856	complement component 4 binding protein, alpha chain precursor	17.34
28	271548	395862	Fibrillin-3; Uncharacterized protein	17.88
28	509129	534490	heterogeneous nuclear ribonucleoprotein M	17.88
28	558963	564318	U6 snRNA-associated Sm-like protein LSm7	17.88
28	564384	599425	signal peptide peptidase-like 2B precursor	17.88
28	627972	628532	translocase of inner mitochondrial membrane 13 homolog (yeast)	17.88
28	636058	661825	lamin-B2	17.88
28	676818	677614	60S ribosomal protein L36	17.88
28	678763	694768	lon peptidase 1, mitochondrial	17.88
28	701484	724473	solute carrier family 1 (high affinity aspartate/glutamate transporter), member 6	17.32
28	899423	901661	anti-dorsalizing morphogenetic protein precursor	17.32
28	932572	950125	ELAV-like protein 1	17.32
28	961655	965629	megakaryocyte-associated tyrosine kinase	17.32
28	972503	976394	Retinal homeobox protein Rx1	17.32
28	976369	983104	mucosa-associated lymphoid tissue lymphoma translocation protein 1-like	17.32
28	991289	1002123	tight junction protein 3	17.32
28	1036061	1041471	cactin, spliceosome C complex subunit	17.32
28	1041856	1045661	thromboxane A2 receptor	17.32
28	1047302	1049378	GIPC PDZ domain containing family, member 3	17.32
28	1049765	1054525	high mobility group 20B	17.32
28	1086890	1089949	Deoxyhypusine hydroxylase	17.32
28	1102700	1127154	unc-13 homolog A (C. elegans)	17.32
28	1143774	1145084	phosphatidic acid phosphatase type 2C	17.32
28	1171971	1179236	amino-terminal enhancer of split	17.32
28	1195931	1202586	guanine nucleotide-binding protein subunit alpha-11	17.32
28	1218202	1225645	nicalin precursor	17.32
28	1233711	1264738	CUGBP, Elav-like family member 5	17.32
28	1266179	1271107	Hydroxysteroid 11-beta-dehydrogenase 1-like protein	17.32
28	1303285	1407097	MPN domain containing	19.72

28	3410986	3425610	solute carrier family 25, member 42	19.72
28	3470619	3483167	homer homolog 3 (Drosophila)	19.72
28	3489321	3495011	probable ATP-dependent RNA helicase DDX49	19.72
28	3495159	3499962	Coatomer subunit epsilon	19.72
28	3514060	3516165	growth differentiation factor 3	19.72
28	3517866	3537861	UPF1 regulator of nonsense transcripts homolog (yeast)	19.72
28	3651445	3653912	cytokine receptor-like factor 1	19.72
28	3657472	3659967	KxDL motif containing 1	19.72
28	3666228	3700318	RNA polymerase II elongation factor ELL	19.72
28	3728960	3735148	LSM4 homolog, U6 small nuclear RNA associated (S. cerevisiae)	19.72
28	3748373	3753211	phosphodiesterase 4C, cAMP-specific	19.72
28	3758472	3759385	MPV17 mitochondrial membrane protein-like 2	19.72
28	3759501	3762347	Interferon-gamma-inducible lysosomal thiol reductase	19.72
28	3782857	3800231	microtubule associated serine/threonine kinase 3	19.72
28	3800995	3805633	interleukin 12 receptor, beta 1	19.72
28	3810197	3816386	arrestin domain containing 2	19.72
28	3820870	3825356	peroxisomal membrane protein 11C	19.72
28	3878507	3912897	Tyrosine-protein kinase receptor	19.72

Supplementary Table 2. List of genes for selective sweeps detected with FLK with 0.05% threshold in white layers.

Chr	Start	End	Description	FLK
1	621303	630454	interferon regulatory factor 5	17.98
1	630567	647745	transportin 3	17.98
1	658443	664773	Smoothened homolog	17.98
1	690901	708802	Adenosylhomocysteinase	17.98
1	710117	727850	striatin interacting protein 2	17.98
1	757267	795092	nuclear respiratory factor 1	17.98
1	807182	849202	ubiquitin-conjugating enzyme E2 H	17.98
1	850460	859444	nuclear-interacting partner of ALK	17.98
1	911097	915397	carboxypeptidase A5 precursor	17.98
1	916281	920045	carboxypeptidase A1 preproprotein	17.98
1	921538	929838	Centrosomal protein of 41 kDa	17.98
1	935413	954122	Coatomer subunit gamma	17.98
1	995787	1011804	ankyrin repeat domain-containing protein 16	17.98
1	1012026	1028924	rab GDP dissociation inhibitor beta	17.98
1	1016281	1061797	family with sequence similarity 208, member B	17.98
1	1068473	1076747	ankyrin repeat and SOCS box protein 13	17.98
1	1088060	1099472	neuroepithelial cell-transforming gene 1 protein	17.98
1	1361642	1822074	exocyst complex component 4	17.98
1	71808627	71817151	BCL2-like 14 (apoptosis facilitator)	17.82
1	71833436	71943987	low density lipoprotein receptor-related protein 6	17.82
1	71974273	71980456	MANSC domain containing 1 precursor	17.82
1	71983303	72042978	Loss of heterozygosity 12 chromosomal region 1 protein homolog	17.82
1	72046586	72113670	dual specificity phosphatase 16	17.82
1	72156810	72166175	G protein-coupled receptor 19	17.82
1	81373015	81656000	Limbic system-associated membrane protein	18.52
1	81680485	81734254	growth associated protein 43	18.52
1	94474405	94635350	glucan (1,4-alpha-), branching enzyme 1	17.82

1	122962850	123267815	FERM and PDZ domain containing 4	17.82
1	123394258	123418544	male-specific lethal 3 homolog (Drosophila)	17.82
2	20547214	20632812	multidrug resistance protein 1	21.02
2	20656627	20694939	RUN domain-containing protein 3B	21.02
2	20696196	20709565	solute carrier family 25, member 40	21.02
2	20711043	20726377	DBF4 homolog (S. cerevisiae)	21.02
2	20865806	20874342	sorcin	21.02
2	30537280	30564128	Sp4 transcription factor	18.90
2	30690282	30698829	cell division cycle-associated 7-like protein	18.90
2	30757166	30909032	Rap guanine nucleotide exchange factor (GEF) 5	18.90
2	30683539	30683661	U5 spliceosomal RNA	18.90
2	37711497	37756128	ubiquitin-conjugating enzyme E2E 1	17.82
2	37769194	37773772	Ribosomal protein L15	17.82
2	37787391	37809859	nuclear receptor subfamily 1 group D member 2	17.82
2	37844468	37965174	Thyroid hormone receptor beta	17.82
2	37698436	37698610	U1 spliceosomal RNA	17.82
2	134048833	134252747	trichorhinophalangeal syndrome I	21.29
2	134628301	134709262	Eukaryotic translation initiation factor 3 subunit H	21.29
3	9262357	9405997	WD repeat containing planar cell polarity effector	17.98
3	9406399	9417141	Malate dehydrogenase, cytoplasmic	17.98
3	9438436	9460691	UTP--glucose-1-phosphate uridylyltransferase	17.98
3	9462114	9489472	vacuolar protein sorting 54 homolog (S. cerevisiae)	17.98
3	9521292	9539167	protein pellino homolog 1	17.72
3	10359052	10449924	Meis homeobox 1	17.72
3	12844082	12878172	Delta-like protein	17.98
3	12887313	12960238	UPF0492 protein C20orf94 homolog	17.98
3	12961577	12970854	McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin	17.98
3	12995964	13027937	Synaptosomal-associated protein 25	17.98
3	16172663	16238767	chromogranin B (secretogranin 1)	17.98
3	16255281	16265499	DNA helicase MCM8	17.98
3	16271778	16276722	cardiolipin synthase 1	17.98
3	16360689	16391192	poly	17.98
3	16404585	16432600	protein lin-9 homolog	17.98
3	16453958	16471806	Golgi resident protein GCP60	17.98
3	16482887	16487420	histone H3.2	17.98
3	16515941	16518570	left-right determination factor 2 precursor	17.98
3	16598307	16603854	signal recognition particle 9 kDa protein	17.78
3	74490287	74540828	mitogen-activated protein kinase kinase kinase 7	17.78
3	74768582	74811042	BTB and CNC homology 1, basic leucine zipper transcription factor 2	17.78
3	74853141	74946896	Midasin	17.78
3	74953170	74985332	ankyrin repeat domain 6	17.78
3	74823486	74824862	gap junction protein, alpha 10, 62kDa	17.78
3	76289510	76312354	synaptotagmin binding, cytoplasmic RNA interacting protein	18.02
3	76326433	76373661	sorting nexin-14	18.02
3	76271065	76271131	Small nucleolar RNA SNORD50	18.02
3	76273207	76273277	Small nucleolar RNA SNORD50	18.02
4	4161930	4176418	solute carrier family 9, subfamily A , member 6	26.70
4	4206227	4209911	four and a half LIM domains 1	26.70
4	4214210	4248185	MAP7 domain containing 3	26.70

4	4280658	4292346	G protein-coupled receptor 112	26.70
4	4301225	4305523	bombesin receptor subtype-3	26.70
4	4345180	4349228	CD40 ligand CD40 ligand, membrane form CD40 ligand, soluble form	26.70
4	4354214	4383568	Rho guanine nucleotide exchange factor 6	26.70
4	4390057	4400409	heterogeneous nuclear ribonucleoprotein G	26.70
4	4398822	4398886	Small nucleolar RNA SNORD61	26.70
4	57404743	57409306	pituitary homeobox 2	21.42
4	57434843	57469617	glutamyl aminopeptidase (aminopeptidase A)	21.42
4	57576632	57597173	Elongation of very long chain fatty acids protein 6	21.42
4	57631764	57683454	pro-epidermal growth factor precursor	21.42
4	57708911	57722892	visual pigment-like receptor peropsin	21.42
4	57731319	57744001	complement factor I	21.42
4	57745870	57751937	phospholipase A2, group XIIA	21.42
4	57754986	57763766	caspase-6	21.42
4	57807176	57923545	PDZ and LIM domain protein 5	17.26
4	82408395	82444435	max dimerization protein 4	17.26
4	82459166	82544586	polymerase (DNA directed) nu	17.26
4	82738354	82761647	negative elongation factor A	17.26
4	82783104	82842064	Wolf-Hirschhorn syndrome candidate 1	26.48
4	89402351	89584157	exocyst complex component 6B	26.48
5	9538613	9582720	LIM domain only 1 (rhombotin 1)	19.31
5	9614317	9626415	resistance to inhibitors of cholinesterase 3 homolog (C. elegans)	19.31
5	9634001	9761468	tubby homolog (mouse)	19.31
5	9768005	9802730	ras-related protein R-Ras2	19.31
5	9834069	9848544	Coatomer subunit beta	19.31
5	9851015	9858153	proteasome subunit alpha type-1	19.31
5	9876942	9951911	cGMP-inhibited 3,5-cyclic phosphodiesterase B	18.02
5	33438355	33483021	sec1 family domain-containing protein 1	18.02
5	33517885	33534657	Cochlin	18.02
5	33543459	33598778	striatin, calmodulin binding protein 3	18.02
5	33616118	33666611	HECT domain containing E3 ubiquitin protein ligase 1	18.02
5	33685490	33736833	HEAT repeat containing 5A	18.02
5	33774090	33856331	nucleotide binding protein-like	18.02
5	34050021	34304671	A kinase (PRKA) anchor protein 6	17.98
5	34375148	34959900	neuronal PAS domain protein 3	17.78
5	44981707	44983006	homeobox protein goosecoid	17.78
5	45092963	45128794	endoribonuclease Dicer	17.78
5	45165017	45184249	calmin (calponin-like, transmembrane)	17.78
5	45270524	45300174	spectrin repeat containing, nuclear envelope family member 3	17.78
5	45330738	45337464	glutaredoxin-related protein 5, mitochondrial	17.78
5	45328658	45328930	Small Cajal body specific RNA 13	17.78
6	3953643	3956579	lung lectin precursor	24.92
6	3961091	3967496	surfactant, pulmonary-associated protein A1 precursor	24.92
6	3981178	3989358	soluble mannose-binding lectin precursor	24.92
6	4283164	4363396	ret proto-oncogene precursor	24.92
6	14667844	14924979	adenosine kinase	26.44
6	14941845	14956153	AP-3 complex subunit mu-1	26.44
6	14961440	14990291	Vinculin	26.44
6	15060198	15068452	urokinase-type plasminogen activator preproprotein	26.44

6	16111040	16115551	zona pellucida sperm-binding protein 4	19.30
6	16273210	16314303	Phosphoinositide 3-kinase adapter protein 1	19.30
6	16318282	16329867	nucleolar and coiled-body phosphoprotein 1	19.30
6	16349917	16354551	ELOVL fatty acid elongase 3	19.30
6	16379333	16476177	golgi brefeldin A resistant guanine nucleotide exchange factor 1	19.30
6	16488268	16493350	nuclear factor NF-kappa-B p100 subunit	19.31
6	24664740	24940362	sortilin-related VPS10 domain containing receptor 1	19.31
6	25264735	25352280	gamma-adducin	19.31
6	30060586	30109050	phosphatidic acid phosphatase type 2 domain containing 1A	19.31
6	30185811	30223053	WD repeat domain 11	19.31
6	30411293	30490016	Fibroblast growth factor receptor 2	19.31
6	30550113	30621520	arginyl-tRNA--protein transferase 1	18.02
6	31636707	31661235	family with sequence similarity 53, member B	18.02
6	31686516	31702105	BRISC complex subunit Abro1	18.02
6	32007895	32012786	testis expressed 36	18.02
6	32058339	32075635	DEAH (Asp-Glu-Ala-His) box polypeptide 32	18.02
7	4312646	4330369	bifunctional purine biosynthesis protein PURH	19.69
7	4333468	4382817	fibronectin precursor	19.69
7	4527554	4590262	NEDD8-conjugating enzyme UBE2F	19.69
7	4591385	4595864	receptor (G protein-coupled) activity modifying protein 1	19.69
7	4611641	4662295	leucine rich repeat (in FLII) interacting protein 1	19.69
7	4722132	4740447	melanophilin	19.69
7	4752831	4805577	collagen alpha-3(VI) chain precursor	19.69
7	11536104	11616253	pleckstrin homology domain containing, family M, member 3	21.24
7	11686741	11704137	cyclic AMP-responsive element-binding protein 1	21.24
7	11766883	11819797	Kruppel-like factor 7 (ubiquitous)	21.24
7	11854607	11868058	carboxypeptidase O	21.24
7	11950838	12015748	disintegrin and metalloproteinase domain-containing protein 23	21.24
7	12072509	12076665	elongation factor 1-beta	21.24
7	12077169	12090936	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	21.24
7	12107101	12129187	INO80 complex subunit D	21.24
7	12074025	12074158	Small nucleolar RNA SNORA41	21.24
7	12074705	12074740	Small nucleolar RNA Z196/R39/R59 family	21.24
7	15219971	15271809	alkylglycerone phosphate synthase	18.46
7	15304546	15320356	nuclear factor erythroid 2-related factor 2	18.46
8	7564281	7592106	smg-7 homolog, nonsense mediated mRNA decay factor	19.69
8	7604691	7609670	actin-related protein 2/3 complex subunit 5	19.69
8	7641619	7708062	ral guanine nucleotide dissociation stimulator-like 1	19.69
8	7903724	7931488	ER degradation enhancer, mannosidase alpha-like 3	19.69
8	7933908	7992849	protein Niban	19.69
8	7564912	7565074	U1 spliceosomal RNA	19.69
9	15087461	15106244	eukaryotic translation initiation factor 4E family member 2	22.97
9	15246615	15264143	eukaryotic translation initiation factor 2B, subunit 5 epsilon, 82kDa	22.97
9	15281328	15292012	dishevelled, dsh homolog 3 (Drosophila)	22.97
9	15293532	15316234	AP-2 complex subunit mu	22.97
9	15332834	15340129	von Willebrand factor A domain containing 5B2	22.97
9	15340601	15343371	ALG3, alpha-1,3- mannosyltransferase	22.97
9	15356178	15362770	endothelin converting enzyme 2	22.97
9	15363725	15369726	26S proteasome non-ATPase regulatory subunit 2	22.97

9	15371232	15387501	eukaryotic translation initiation factor 4 gamma, 1	22.97
9	15395572	15405268	chloride channel, voltage-sensitive 2	22.97
9	15406844	15413438	chordin precursor	22.97
9	15419502	15421617	thrombopoietin precursor	22.97
9	15437757	15449465	EPH receptor B3	22.97
9	15380309	15380385	Small nucleolar RNA SNORD66	22.97
9	15381519	15381596	Small nucleolar RNA SNORD66	22.97
9	16565596	16587667	fragile X mental retardation syndrome-related protein 1	17.61
9	16663283	16672397	tetratricopeptide repeat protein 14	17.61
9	16760382	16850020	peroxisomal biogenesis factor 5-like	17.61
9	16861502	16902461	Ubiquitin carboxyl-terminal hydrolase 13	17.61
9	16903953	16907474	NADH dehydrogenase	17.61
9	16911951	16921482	actin-like 6A	17.61
9	16984938	17003984	mitofusin-1	17.61
9	16820695	16820798	U6 spliceosomal RNA	17.61
9	20419761	20464371	cholinesterase precursor	17.72
9	20586079	20587887	SLIT and NTRK-like family, member 3	17.72
10	6434758	6619897	disintegrin and metalloproteinase domain-containing protein 10 precursor	22.97
10	6572777	6596733	aquaporin 9	22.97
10	6604818	6661741	Retinal dehydrogenase 2	22.97
10	6816944	6857906	cingulin-like 1	22.97
10	6903270	7058903	transcription factor 12	18.85
10	9545024	9553009	solute carrier family 24, member 5 precursor	18.85
10	9551752	9570906	myelin expression factor 2	18.85
10	9587065	9632792	solute carrier family 12 (sodium/potassium/chloride transporters), member 1	18.85
10	17122370	17146830	proprotein convertase subtilisin/kexin type 6	17.82
10	17199826	17220109	ankyrin repeat and death domain containing 1A	17.83
10	18440729	18553284	dual specificity mitogen-activated protein kinase kinase 5	17.83
10	18595984	18624673	E3 SUMO-protein ligase PIAS1	17.83
10	18632517	18638897	ceroid-lipofuscinosis, neuronal 6, late infantile, variant	17.83
10	18641302	18648029	protein fem-1 homolog B	17.83
10	18699925	18729841	coronin, actin binding protein, 2B	17.83
10	18770495	18779187	NADPH oxidase 5	17.83
10	18807760	18817356	glucuronic acid epimerase	17.83
10	18823404	18826891	progestin and adipoQ receptor family member V	17.83
10	18828578	18846365	kinesin-like protein KIF23	17.83
11	983068	995051	E2F transcription factor 4; Uncharacterized protein	21.24
11	995632	1005522	engulfment and cell motility 3	21.24
11	1011955	1025358	KIAA0895-like	21.24
11	1029318	1036713	lecithin-cholesterol acyltransferase	21.24
11	1073882	1097277	family with sequence similarity 65, member A	21.24
11	1102641	1128402	transcriptional repressor CTCF	21.24
11	1130092	1147360	RGD motif, leucine rich repeats, tropomodulin domain and proline-rich containing	21.24
11	1165618	1177771	alanyl-tRNA synthetase, cytoplasmic	21.24
11	1189504	1194020	fibulin 7	21.24
11	1213866	1215689	tubulin polymerization-promoting protein family member 3	21.24
11	1223848	1237337	potassium channel tetramerisation domain containing 19	21.24
11	1240508	1302916	pleckstrin homology domain containing, family G member 4	21.24
11	1336785	1351404	FH1/FH2 domain-containing protein 1	21.24

11	1389699	1391495	agouti-related protein precursor	21.24
11	1398993	1402340	N-lysine methyltransferase SETD6	21.24
11	1405351	1453089	CCR4-NOT transcription complex, subunit 1	21.24
11	1424514	1424649	Small nucleolar RNA SNORA46	21.24
11	1432730	1432863	Small nucleolar RNA SNORA76	21.24
11	7596758	7603866	Cytochrome b-c1 complex subunit Rieske, mitochondrial	17.72
11	7841312	7844172	processing of precursor 4	17.72
11	7933476	7946064	G1/S-specific cyclin-E1	17.72
11	7961122	8028603	URI1, prefoldin-like chaperone	17.72
13	2010626	2012530	leucine rich repeat transmembrane neuronal 2	30.29
13	2134785	2156387	stress-70 protein, mitochondrial precursor	30.29
13	2146299	2146368	Small nucleolar RNA SNORD63	30.29
14	5075181	5086183	DNA topoisomerase 3-alpha	19.43
14	5112713	5127320	phosphoribosyl pyrophosphate synthase-associated protein 2	19.43
14	5129153	5157900	solute carrier family 5 (sodium/glucose cotransporter), member 10	19.43
14	5138381	5150303	family with sequence similarity 83, member G	19.43
14	5223282	5225616	B9 protein domain 1	19.43
14	5230050	5338554	Voltage-dependent T-type calcium channel subunit alpha-1H	19.43
14	9934515	9974931	ubiquitin carboxyl-terminal hydrolase 7	22.75
14	10046044	10052775	phosphomannomutase 2	22.75
14	10103934	10114456	methyltransferase-like protein 22	22.75
17	4788861	4790491	ovoglycoprotein precursor	17.26
17	4820725	4824113	nuclear apoptosis inducing factor 1	17.26
17	4829664	4860591	early estrogen-induced gene 1 protein	17.26
17	4865441	4867919	dolichol phosphate-mannose biosynthesis regulatory protein	17.26
17	4870177	4871635	alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1	17.26
17	4872095	4874530	ST6 -N-acetylgalactosaminide alpha-2,6-sialyltransferase 6	17.26
17	4877410	4880796	Adenylate kinase isoenzyme 1	17.26
17	4882932	4891674	endoglin precursor	17.26
17	4892549	4896134	folylpolyglutamate synthase	17.26
17	4896492	4902733	Cyclin-dependent kinase 9	17.26
17	4903795	4923543	SH2 domain containing 3C	17.26
17	4929079	4931968	tetratricopeptide repeat domain 16	17.26
17	4933496	4946785	cerebral endothelial cell adhesion molecule	17.26
17	4949346	4967354	Ubiquitin-related modifier 1 homolog	17.26
17	4983785	4987536	coenzyme Q4 homolog (S. cerevisiae)	17.26
17	5016759	5075578	dynamin 1	17.82
18	2020996	2066954	Netrin-1	17.82
18	2070106	2146328	syntaxin-8	17.82
18	2180734	2183708	ubiquitin specific peptidase 43	17.82
18	2188691	2205886	Rho GTPase activating protein 44	17.82
18	2223389	2230202	uncharacterized protein LOC417324	17.82
19	7951502	7972199	protein phosphatase, Mg2+/Mn2+ dependent, 1D	19.69
19	7979200	7997166	amyloid protein-binding protein 2	19.69
19	8096015	8104574	carbonic anhydrase IV	19.69
19	8148378	8165006	Gametogenetin-binding protein 2	19.69
19	8166412	8187734	Dehydrogenase/reductase SDR family member 11	19.69
21	3252441	3261217	G protein-coupled receptor 157	19.69
21	3470173	3492933	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic	19.69

21	3493075	3527740	calsyntenin-1 precursor	19.69
21	3573056	3581801	Protein LZIC	19.69
21	3581913	3588648	nicotinamide nucleotide adenylyltransferase 1	19.69
21	3590985	3592942	retinol binding protein 7, cellular	19.69
21	3595756	3630094	ubiquitination factor E4B	19.69
21	3639974	3713856	kinesin family member 1B	19.69
21	3719439	3729173	6-phosphogluconate dehydrogenase, decarboxylating	19.69
21	3730260	3733060	Centromere protein S	19.69
21	3705900	3706069	TUC338	19.69
22	1256816	1265197	Charged multivesicular body protein 7	19.43
22	1270094	1276295	tumor necrosis factor receptor superfamily member 10B precursor	19.43
22	1343627	1375329	exportin-7	19.43
22	1376680	1381157	docking protein 2, 56kDa	19.43
22	1397665	1418597	GDNF family receptor alpha-2 precursor	19.43
22	1597651	1602693	TELO2 interacting protein 2	19.43
23	4570631	4804042	CUB and Sushi multiple domains 2	17.82
23	4824370	4834663	collagen, type IX, alpha 2	17.82
23	4835192	4847417	small ArfGAP2	17.82
23	4901524	4907272	potassium voltage-gated channel, KQT-like subfamily, member 4	17.82
23	4910836	4916448	tubulointerstitial nephritis antigen-like 1	17.82
23	4918792	4920966	penta-EF-hand domain containing 1	17.82
23	4921631	4939423	collagen, type XVI, alpha 1	17.82
23	4943828	4960852	brain-specific angiogenesis inhibitor 2	17.82
24	2427011	2464846	immunoglobulin superfamily, member 9B	19.60
24	2516848	2542717	non-SMC condensin II complex, subunit D3	19.60
24	2553386	2558481	Thymocyte nuclear protein 1	19.60
24	2845500	2863118	potassium-transporting ATPase alpha chain 2	19.60
24	2902288	3002638	GRAM domain containing 1B	22.75
24	4392700	4566947	ubiquitin conjugation factor E4 A	22.75
24	4557911	4561306	RNA-binding protein 7	22.75
24	4664469	4732605	cell adhesion molecule 1	22.75
26	2504099	2506388	Interleukin-10	17.01
26	2517673	2519712	interleukin 19	17.01
26	2520032	2531914	polymeric immunoglobulin receptor precursor	17.01
26	2556703	2560172	ubiquitin thioesterase OTU1	17.01
26	2576519	2583374	C4b-binding protein alpha chain precursor	17.01
26	2599568	2605277	complement component (3b/4b) receptor 1-like precursor	17.01
26	2627129	2639856	complement component 4 binding protein, alpha chain precursor	17.01
28	271548	395862	Fibrillin-3; Uncharacterized protein	17.61
28	509129	534490	heterogeneous nuclear ribonucleoprotein M	17.61
28	558963	564318	U6 snRNA-associated Sm-like protein LSm7	17.61
28	564384	599425	signal peptide peptidase-like 2B precursor	17.61
28	627972	628532	translocase of inner mitochondrial membrane 13 homolog (yeast)	17.61
28	636058	661825	lamin-B2	17.61
28	676818	677614	60S ribosomal protein L36	17.61
28	678763	694768	lon peptidase 1, mitochondrial	17.61
28	701484	724473	solute carrier family 1, member 6	19.43
28	3410986	3425610	solute carrier family 25, member 42	19.43
28	3470619	3483167	homer homolog 3 (Drosophila)	19.43

28	3489321	3495011	probable ATP-dependent RNA helicase DDX49	19.43
28	3495159	3499962	Coatomer subunit epsilon	19.43
28	3514060	3516165	growth differentiation factor 3	19.43
28	3517866	3537861	UPF1 regulator of nonsense transcripts homolog (yeast)	19.43
28	3651445	3653912	cytokine receptor-like factor 1	19.43
28	3657472	3659967	KxDL motif containing 1	19.43
28	3666228	3700318	RNA polymerase II elongation factor ELL	19.43
28	3728960	3735148	LSM4 homolog, U6 small nuclear RNA associated (S. cerevisiae)	19.43
28	3748373	3753211	phosphodiesterase 4C, cAMP-specific	19.43
28	3758472	3759385	MPV17 mitochondrial membrane protein-like 2	19.43
28	3759501	3762347	Interferon-gamma-inducible lysosomal thiol reductase	19.43
28	3782857	3800231	microtubule associated serine/threonine kinase 3	19.43
28	3800995	3805633	interleukin 12 receptor, beta 1	19.43
28	3810197	3816386	arrestin domain containing 2	19.43
28	3820870	3825356	peroxisomal membrane protein 11C	19.43
28	3878507	3912897	Tyrosine-protein kinase receptor	19.43

Supplementary Table 3. List of genes for selective sweeps detected with FLK with 0.05% threshold in brown layers.

Chr	Start	End	Description	FLK
1	38593310	38850635	neuron navigator 3	15.29
1	110613430	110658911	monoamine oxidase A	15.29
2	77784558	78033401	catenin (cadherin-associated protein), delta 2	14.36
3	29337956	29424515	MAM domain-containing glycosylphosphatidylinositol anchor protein 1 precursor	14.36
3	29217044	29217179	TUC338	14.36
4	38695965	38707538	ufm1-specific protease 2	14.36
4	38707911	38715236	LRP2 binding protein	14.36
4	38717204	38769907	sorting nexin 25	14.36
4	38786566	38804359	KIAA1430	14.36
4	48659743	48674369	signal recognition particle 72 kDa protein	14.36
4	48676608	48680544	ADP-ribosylation factor-like 9	14.76
4	48685349	48688644	Homeodomain-only protein	14.76
4	48706786	48712370	serine peptidase inhibitor, Kazal type 2 (acrosin-trypsin inhibitor)	14.76
4	48787273	48788448	dual specificity protein phosphatase 4	14.76
4	48794273	48900800	tankyrase-1	14.76
4	48927640	48936506	3-5 exoribonuclease 1	14.76
4	48960039	48966650	Protein phosphatase 1 regulatory subunit 3	14.76
4	49033103	49034980	ghrelin O-acyltransferase	14.76
4	49041665	49042706	Ribonuclease CL2	14.76
4	49077854	49125067	septin 11	14.76
4	49129165	49130967	sosondowah ankyrin repeat domain family member B	14.76
4	49142913	49148727	16 kDa beta-galactoside-binding lectin	14.76
4	49156409	49291621	shroom family member 3	14.76
5	30717442	30832671	uncharacterized protein C15orf41 homolog	16.23
11	5636333	5654522	cylindromatosis (turban tumor syndrome)	16.23
11	5725577	5829415	naked cuticle homolog 1 (Drosophila)	16.23
11	5990712	6006669	bromodomain-containing protein 7	16.23
11	6010767	6037599	adenylate cyclase 7	16.23
13	185486	285392	protocadherin alpha 11 precursor	16.23

13	365453	377641	histidyl-tRNA synthetase, cytoplasmic	16.23
13	384539	392849	protein Red	16.23
13	392987	394389	NADH dehydrogenase	16.23
13	418871	424514	Transmembrane protein 173	16.23
13	402430	403827	CD14 molecule precursor	16.23
13	380747	380840	Vault RNA	16.23
18	9623865	9743508	regulatory associated protein of MTOR, complex 1	15.29
18	9754645	9756687	neuronal pentraxin I	15.29
18	9763404	9771339	endonuclease V	15.29
18	9771940	9818172	ring finger protein 213	15.29
18	9819301	9824829	solute carrier family 26, member 11	15.29
18	9885036	9886813	E3 SUMO-protein ligase CBX4	15.29
18	10016690	10022596	RNA binding protein, fox-1 homolog (C. elegans) 3	15.29
18	10024650	10029994	endo-beta-N-acetylglucosaminidase	15.29
18	10034762	10038434	soluble calcium-activated nucleotidase 1	15.29
18	10040657	10047611	metalloproteinase inhibitor 2 precursor	15.29
18	10062346	10070304	cytohesin-1	15.29
18	10070951	10098653	dynein, axonemal, heavy chain 17	15.29
18	10096967	10115046	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	15.29
18	10142947	10144362	Thymidine kinase, cytosolic	15.29
18	10145914	10147671	synaptogyrin 2	15.29
18	10154424	10156574	ADP-ribosylation factor-like 16	15.29
18	10181308	10184603	gastric inhibitory polypeptide receptor precursor	15.29
18	10183770	10202043	protein disulfide-isomerase precursor	15.29
18	10116718	10117433	suppressor of cytokine signaling 3	15.29
21	3826146	3869802	castor zinc finger 1	16.23
21	4065712	4071374	TAR DNA-binding protein 43	16.23
21	4073317	4086699	mannan-binding lectin serine protease 2 precursor	16.23
21	4099857	4103582	Pro2-somatostatin precursor	16.23
21	4118177	4132290	exosome component 10	16.23
21	4133269	4193503	mechanistic target of rapamycin (serine/threonine kinase)	16.23
21	4195215	4200725	ubiA prenyltransferase domain-containing protein 1	16.23
21	4202008	4221119	pleckstrin homology domain containing, family M member 2	16.23
21	4229458	4231641	filamin binding LIM protein 1	16.23
21	4260524	4301058	spen homolog, transcriptional regulator (Drosophila)	16.23
21	4191831	4192009	TUC338	16.23
23	3965824	3970617	granulocyte colony-stimulating factor receptor	14.44
23	3977706	3987124	organic solute carrier partner 1	14.44
23	3988933	4004007	serine/threonine-protein kinase 40	14.44
23	4007214	4007707	eva-1 homolog B (C. elegans)	14.44
23	4008688	4024899	thyroid hormone receptor-associated protein 3	14.44
23	4031627	4038165	MAP7 domain containing 1	14.44
23	4050995	4056202	Trafficking protein particle complex subunit 3	14.44
23	4092973	4097739	poly(ADP-ribose) glycohydrolase ARH3	14.44
23	4098117	4102886	tektin 2 (testicular)	14.44
23	4111971	4133770	protein argonaute-3	14.44
23	4176483	4190009	claspin	14.44
23	4222202	4234872	Proteasome subunit beta type	14.44
23	4256435	4263091	neurochondrin	14.44

23	4264511	4288627	KIAA0319-like	14.44
23	4295102	4296014	interferon alpha-inducible protein 27-like protein 2	14.44
23	4343978	4352202	Gizzard PTB-associated splicing factor; Uncharacterized protein	14.44
23	4361390	4361969	ZMYM6 neighbor	14.44
23	4379948	4393033	discs, large (Drosophila) homolog-associated protein 3	14.44
23	4398099	4400981	connexin 37	14.44
23	4404012	4404779	gap junction protein, beta 3, 31kDa	14.44
27	2226991	2253136	corticotropin-releasing factor receptor 1 precursor	14.19
27	2302640	2319626	integrin beta-3 precursor	14.19
27	2332100	2344617	Methyltransferase-like protein 2	14.19
27	2357455	2399073	tousled-like kinase 2	14.19
27	2412671	2426010	mannose receptor, C type 2	14.19
27	2524612	2629807	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2	14.19
27	2634670	2639226	cytochrome b561	14.19
27	3204183	3219385	tumor necrosis factor receptor superfamily member 16 precursor	14.19
27	3252521	3267354	membrane protein, palmitoylated 3	14.76
27	3315353	3321406	homeobox protein MOX-1	14.76
27	3335561	3350183	ets variant 4	14.76
27	3352213	3364975	DEAH (Asp-Glu-Ala-His) box polypeptide 8	14.76
27	3366784	3370919	Prohibitin	14.76
27	3413412	3417057	phosphoethanolamine/phosphocholine phosphatase	14.76
27	3423443	3425719	guanine nucleotide binding protein, gamma transducing activity polypeptide 2	14.76
27	3433057	3459259	Insulin-like growth factor 2 mRNA-binding protein 1	14.76
27	3468107	3475761	gastric inhibitory polypeptide precursor	14.76
27	3495909	3506537	calcium binding and coiled-coil domain 2	14.76
27	3586132	3589844	Hoxb-7	14.76
27	3598354	3600970	homeobox B6	14.76
27	3604171	3606516	Homeobox protein Hox-B5	14.76
27	3621645	3626317	homeobox protein Hox-B4	14.76
27	3643412	3649970	homeobox protein Hox-B3	14.76
27	3652538	3655356	homeobox B2	14.76
27	3662963	3664600	homeobox B1	14.76
27	3711055	3815879	src kinase associated phosphoprotein 1	14.76
28	3410986	3425610	solute carrier family 25, member 42	14.36
28	3470619	3483167	homer homolog 3 (Drosophila)	14.36
28	3489321	3495011	probable ATP-dependent RNA helicase DDX49	14.36
28	3495159	3499962	Coatomer subunit epsilon	14.36
28	3514060	3516165	growth differentiation factor 3	14.36
28	3517866	3537861	UPF1 regulator of nonsense transcripts homolog (yeast)	14.36
28	3651445	3653912	cytokine receptor-like factor 1	14.36
28	3657472	3659967	KxDL motif containing 1	14.36
28	3666228	3700318	RNA polymerase II elongation factor ELL	14.36
28	3728960	3735148	LSM4 homolog, U6 small nuclear RNA associated	14.36
28	3748373	3753211	phosphodiesterase 4C, cAMP-specific	14.36
28	3758472	3759385	MPV17 mitochondrial membrane protein-like 2	14.36
28	3759501	3762347	Interferon-gamma-inducible lysosomal thiol reductase	14.36
28	3782857	3800231	microtubule associated serine/threonine kinase 3	14.36
28	3800995	3805633	interleukin 12 receptor, beta 1	14.36
28	3810197	3816386	arrestin domain containing 2	14.36

28	3820870	3825356	peroxisomal membrane protein 11C	14.36
28	4195685	4203208	tropomyosin 4	15.29
28	4410520	4497158	receptor-type tyrosine-protein phosphatase delta precursor	15.29
28	4512762	4588636	lysine (K)-specific demethylase 4B	15.29
28	4673752	4692777	dipeptidyl-peptidase 9	15.29
28	4625272	4627738	toll-like receptor adaptor molecule 1	15.29

Supplementary Table 4. Lists of pathways and gene ontologies under selection with FLK with 0.05% threshold in all layers.

Description	**# Genes anotated**	**Genes of pathways (%)***	**P-Value**
membrane	22	1.1	0.000
integral to membrane	21	1.3	0.000
zinc ion binding	18	1.5	0.000
intracellular	18	1.5	0.000
plasma membrane	15	1.5	0.000
Metabolic pathways	17	12.5	0.000
ATP binding	19	1.8	0.000
DNA binding	16	1.7	0.000
nucleotide binding	17	1.8	0.000
mitochondrion	17	1.9	0.000
regulation of transcription, DNA-dependent	15	1.9	0.000
signal transduction	17	2.2	0.000
metal ion binding	14	2.0	0.001
COPI-coated vesicle membrane	3	60.0	0.001
maintenance of DNA methylation	3	60.0	0.001
nucleic acid binding	13	2.1	0.002
intercellular bridge	3	50.0	0.002
protein K63-linked deubiquitination	4	30.8	0.002
cellular response to oxidative stress	4	28.6	0.003
positive regulation of transcription from RNA polymerase II	9	2.0	0.004
G-protein coupled receptor signaling pathway	8	1.9	0.004
COPI vesicle coat	3	37.5	0.005
exocyst	3	37.5	0.005
regulation of cell size	3	37.5	0.005
endoplasmic reticulum unfolded protein response	4	23.5	0.006
urogenital system development	3	30.0	0.009
sequence-specific DNA binding transcription factor activity	13	2.4	0.009
ubiquitin ligase complex	5	16.7	0.011
retinoic acid receptor signaling pathway	3	27.3	0.012
Neuroactive ligand-receptor interaction	5	12.5	0.012
protein phosphorylation	12	2.4	0.013
regulation of cyclin-dependent protein serine/threonine kinase activity	4	17.4	0.019
vesicle docking involved in exocytosis	3	23.1	0.019
cytosol	9	2.2	0.021
sequence-specific DNA binding	11	2.4	0.022
ubiquitin-specific protease activity	5	13.9	0.023
dorsal/ventral neural tube patterning	3	21.4	0.024
Golgi apparatus	10	2.4	0.024

transferase activity, transferring phosphorus-containing groups	12	2.5	0.026
positive regulation of proteolysis	3	20.0	0.029
endoplasmic reticulum	8	2.2	0.029
cell periphery	4	14.8	0.033
one-carbon metabolic process	3	18.8	0.034
ATP metabolic process	3	18.8	0.034
guanyl-nucleotide exchange factor activity	6	11.1	0.036
signal transducer activity	8	2.3	0.038
G-protein coupled receptor activity	8	2.3	0.038
negative regulation of intrinsic apoptotic signaling pathway	3	17.6	0.040
negative regulation of tumor necrosis factor production	3	17.6	0.040
One carbon pool by folate	3	12.5	0.040
Galactose metabolism	4	12.5	0.040
protein kinase activity	12	2.6	0.041
negative regulation of NF-kappaB transcription factor activity	4	13.8	0.042
transcription, DNA-dependent	6	2.1	0.044
extracellular region	11	2.6	0.046
hair follicle morphogenesis	3	16.7	0.047
positive regulation of fat cell differentiation	3	16.7	0.047
G2 DNA damage checkpoint	3	16.7	0.047
ubiquitin thiolesterase activity	6	10.3	0.049

*Percentage of the genes of the pathway which were among the annotated genes.

Supplementary Table 5. Lists of pathways and gene ontologies under selection with FLK with 0.05% threshold in white layers.

Description	# Genes anotated	Genes of pathways (%)*	P-Value
membrane	21	1.1	0.000
integral to membrane	20	1.3	0.000
zinc ion binding	17	1.4	0.000
Metabolic pathways	16	33.3	0.000
intracellular	18	1.5	0.000
plasma membrane	14	1.4	0.000
ATP binding	17	1.6	0.000
DNA binding	16	1.7	0.000
mitochondrion	16	1.8	0.000
nucleotide binding	17	1.8	0.000
COPI-coated vesicle membrane	3	60.0	0.001
maintenance of DNA methylation	3	60.0	0.001
intercellular bridge	3	50.0	0.001
metal ion binding	12	1.7	0.001
regulation of transcription, DNA-dependent	15	1.9	0.001
COPI vesicle coat	3	37.5	0.003
exocyst	3	37.5	0.003
regulation of cell size	3	37.5	0.003
endoplasmic reticulum unfolded protein response	4	23.5	0.004
signal transduction	16	2.1	0.004
urogenital system development	3	30.0	0.006
nucleic acid binding	12	2.0	0.008
retinoic acid receptor signaling pathway	3	27.3	0.008

protein phosphorylation	9	1.8	0.010
Neuroactive ligand-receptor interaction	4	33.3	0.013
vesicle docking involved in exocytosis	3	23.1	0.013
protein K63-linked deubiquitination	3	23.1	0.013
dorsal/ventral neural tube patterning	3	21.4	0.016
cellular response to oxidative stress	3	21.4	0.016
transferase activity, transferring phosphorus-containing	9	1.9	0.016
guanyl-nucleotide exchange factor activity	6	11.1	0.019
cell periphery	4	14.8	0.020
G-protein coupled receptor signaling pathway	8	1.9	0.022
one-carbon metabolic process	3	18.8	0.023
ATP metabolic process	3	18.8	0.023
positive regulation of transcription from RNA polymerase II	9	2.0	0.027
protein kinase activity	9	1.9	0.027
transcription, DNA-dependent	4	1.4	0.028
ubiquitin ligase complex	4	13.3	0.029
hair follicle morphogenesis	3	16.7	0.032
positive regulation of fat cell differentiation	3	16.7	0.032
heart looping	4	12.9	0.032
regulation of pH	3	15.8	0.036
protein tyrosine kinase activity	8	2.0	0.038
oxidoreductase activity	5	10.4	0.039
metallocarboxypeptidase activity	3	15.0	0.042
heart morphogenesis	4	11.8	0.043

*Percentage of the genes of the pathway which were among the annotated genes.

Supplementary Table 6. Lists of pathways and gene ontologies under selection with FLK with 0.05% threshold in brown layers.

Description	# Genes anotated	Genes of pathways (%)*	P-Value
histone mRNA catabolic process	3	33.3	0.000
phosphoprotein binding	3	13.6	0.003
peptidyl-serine phosphorylation	4	7.8	0.006
stress fiber	3	7.7	0.017
mTOR signaling pathway	3	10.0	0.029
Metabolic pathways	6	12.1	0.033

*Percentage of the genes of the pathway which were among the annotated genes.

Supplementary Table 7. List of genes for selective sweeps detected with hapFLK with 0.05% threshold in all layers.

Chr	Start	End	Description	hapFLK
1	50719186	50733066	DMC1 dosage suppressor of mck1 homolog, meiosis-specific homologous	7.91
1	50753266	50759412	KDEL receptor 3	7.91
1	50770322	50773091	Inward-rectifying potassium channel cKir2.3; Uncharacterized protein	7.91
1	50788491	50804577	casein kinase I isoform epsilon	7.91
1	50832960	50835917	Transcription factor MafF	7.91
1	50836439	50854357	85 kDa calcium-independent phospholipase A2	7.91
1	50856306	50864314	BAI1-associated protein 2-like 2	7.91
1	50865712	50873377	Monocarboxylate transporter 3	7.91

1	50911765	50921215	Transcription factor SOX-10	7.91
1	50924002	50927570	DNA-directed RNA polymerases I, II, and III subunit RPABC2	7.91
1	50931992	50950143	MICAL-like protein 1	7.91
1	50953763	50963376	Eukaryotic translation initiation factor 3 subunit L	7.91
1	50979163	50980997	galanin receptor type 3	7.91
1	51010961	51014093	Beta-galactoside-binding lectin	7.91
1	51026210	51029611	pyridoxal (pyridoxine, vitamin B6) phosphatase	7.91
1	51050593	51062352	lectin, galactoside-binding, soluble, 2	7.91
1	51065196	51067106	CDC42 effector protein (Rho GTPase binding) 1	7.91
1	51083236	51095519	caspase recruitment domain family, member 10	7.91
1	51099652	51110327	MFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	7.91
1	51229151	51248424	cytohesin-4	7.91
1	51257450	51266551	ras-related C3 botulinum toxin substrate 2	7.91
1	51269724	51274129	somatostatin receptor type 3	7.91
1	51336790	51344019	potassium channel tetramerisation domain containing 17	7.91
1	51349989	51355027	Sulfurtransferase	7.91
1	51355166	51362344	thiosulfate sulfurtransferase	7.91
1	51419978	51430209	neutrophil cytosol factor 4	7.91
1	51455040	51463393	Parvalbumin, muscle	7.91
1	51573280	51580672	thioredoxin, mitochondrial	7.91
1	51598550	51666512	myosin-9	7.91
1	50983576	50984295	noggin 4 precursor	7.91
1	50991976	50992548	Histone H5	7.91
1	51511551	51511715	TUC338	7.91
1	5715214	5839991	CUGBP Elav-like family member 2	1.79
1	8177104	8318907	Semaphorin-3D	1.79
1	8522175	8850302	semaphorin-3A	1.79
1	127508430	127539291	protein kinase, X-linked	1.81
1	127679910	127698770	matrix-remodelling associated 5	1.81
1	127805905	127818492	arylsulfatase H precursor	1.81
1	127871298	127881303	glycogenin 2	1.81
1	127916919	127941663	CD99 antigen precursor	1.81
2	134048833	134252747	trichorhinophalangeal syndrome I	7.85
2	134628301	134709262	Eukaryotic translation initiation factor 3 subunit H	7.85
2	134749874	134773800	double-strand-break repair protein rad21 homolog	7.85
2	11735794	11743710	Krueppel-like factor 6	1.81
2	73536943	73537003	Small nucleolar RNA R11/Z151	1.81
2	113502155	113517713	tocopherol (alpha) transfer protein	1.81
2	113543309	113551391	YTH domain family protein 3	1.81
2	114472659	114503139	Armadillo repeat-containing protein 1	1.81
2	113888475	113888659	TUC338	1.81
2	142700486	142776708	KH domain containing, RNA binding, signal transduction associated 3	1.81
3	6450555	7094776	neurexin-1-alpha isoform 1 precursor	1.81
3	34322339	34455636	v-akt murine thymoma viral oncogene homolog 3	1.81
3	34464087	34563969	serologically defined colon cancer antigen 8	1.81
3	34813540	34977620	inactive phospholipase D5	1.81
3	35065030	35080589	exonuclease 1	1.81
3	35088960	35147574	WD repeat domain 64	1.81
3	35238582	35471485	regulator of G-protein signaling 7	1.81

3	35212222	35213478	GPI mannosyltransferase 1	1.81
6	14411231	14417212	voltage-dependent anion-selective channel protein 2	8.06
6	14522141	14541652	Dual specificity phosphatase DUPD1	8.06
6	14667844	14924979	adenosine kinase	8.06
6	14941845	14956153	AP-3 complex subunit mu-1	8.06
6	14961440	14990291	Vinculin	8.06
6	15060198	15068452	urokinase-type plasminogen activator preproprotein	8.06
13	1767729	1785069	endothelial cell surface expressed chemotaxis and apoptosis regulator	8.38
13	1788284	1799938	dnaJ homolog subfamily C member 18	8.38
13	1806725	1808111	marginal zone B and B1 cell-specific protein	8.38
13	1813947	1819364	Polyadenylate-binding protein-interacting protein 2	8.38
13	1825981	1851149	matrin-3	8.38
13	2010626	2012530	leucine rich repeat transmembrane neuronal 2	8.38
13	2134785	2156387	stress-70 protein, mitochondrial precursor	8.38
13	2567435	2611832	fibroblast growth factor 18 precursor	8.38
13	2641231	2651587	nucleophosmin	8.38
13	2681554	2828621	RAN binding protein 17	8.38
13	2146299	2146368	Small nucleolar RNA SNORD63	8.38
13	10000421	10264193	transcription factor COE1	1.81
13	10554868	10603091	clathrin interactor 1	1.81
13	10618933	10623222	LSM11, U7 small nuclear RNA associated	1.81
13	10624187	10630514	probable tRNA(His) guanylyltransferase	1.81
13	10642122	10669352	disintegrin and metalloproteinase domain-containing protein 19 precursor	1.81
13	10668257	10680834	NIPA-like domain containing 4	1.81
13	10737034	10764716	IL2-inducible T-cell kinase	1.81
13	10778915	10785137	hepatitis A virus cellular receptor 1 precursor	1.81
27	3080955	3092426	histone acetyltransferase MYST2	8.98
27	3121453	3124525	solute carrier family 35 member B1	8.98
27	3204183	3219385	tumor necrosis factor receptor superfamily member 16 precursor	8.98
27	3252521	3267354	membrane protein, palmitoylated 3	8.98
27	3315353	3321406	homeobox protein MOX-1	8.98
27	3335561	3350183	ets variant 4	8.98
27	3352213	3364975	DEAH (Asp-Glu-Ala-His) box polypeptide 8	8.98
27	3366784	3370919	Prohibitin	8.98
27	3413412	3417057	phosphoethanolamine/phosphocholine phosphatase	8.98
27	3423443	3425719	guanine nucleotide binding protein , gamma transducing activity polypeptide 2	8.98
27	3433057	3459259	Insulin-like growth factor 2 mRNA-binding protein 1	8.98
27	3468107	3475761	gastric inhibitory polypeptide precursor	8.98
27	3495909	3506537	calcium binding and coiled-coil domain 2	8.98
27	3586132	3589844	Hoxb-7	8.98
27	3598354	3600970	homeobox B6	8.98
27	3604171	3606516	Homeobox protein Hox-B5	8.98
27	3621645	3626317	homeobox protein Hox-B4	8.98
27	3643412	3649970	homeobox protein Hox-B3	8.98
27	3652538	3655356	homeobox B2	8.98
27	3662963	3664600	homeobox B1	8.98
27	3711055	3815879	src kinase associated phosphoprotein 1	8.98
27	3842373	3850380	chromobox protein homolog 1	8.98
27	3855203	3861688	Nuclear factor erythroid 2-related factor 1	8.98

27	3865615	3868783	CDK5 regulatory subunit associated protein 3	8.98
27	3896638	3899664	leucine rich repeat containing 46	8.98
27	3902862	3907297	oxysterol binding protein-like 7	8.98
27	3971094	4005910	myeloid/lymphoid or mixed-lineage leukemia ; translocated to, 6	8.98
27	4006813	4007489	CDGSH iron sulfur domain 3	8.98
27	4008146	4011942	polycomb group ring finger 2	8.98
27	4014223	4017030	Proteasome subunit beta type	8.98
27	4046906	4048844	ribosomal protein L23	8.98
27	4053819	4068552	LIM and SH3 domain protein 1	8.98
27	3869421	3870521	proline rich 15-like	8.98
27	3155867	3156043	TUC338	8.98
27	4047757	4047888	Small nucleolar RNA SNORA21	8.98
28	627972	628532	translocase of inner mitochondrial membrane 13 homolog (yeast)	7.93
28	636058	661825	lamin-B2	7.93
28	676818	677614	60S ribosomal protein L36	7.93
28	678763	694768	lon peptidase 1, mitochondrial	7.93
28	701484	724473	solute carrier family 1 (high affinity aspartate/glutamate transporter), member 6	7.93
28	746840	778361	RAN binding protein 3	7.93
28	789271	792789	kelch-like family member 33	7.93
28	803320	808851	butyrophilin subfamily 1 member A1 precursor	7.93
28	832292	846222	ras-related protein Rab-11B	7.93
28	877452	886036	KN motif and ankyrin repeat domains 3	7.93
28	899423	901661	anti-dorsalizing morphogenetic protein precursor	7.93
28	932572	950125	ELAV-like protein 1	7.93
28	961655	965629	megakaryocyte-associated tyrosine kinase	7.93
28	972503	976394	Retinal homeobox protein Rx1	7.93
28	976369	983104	mucosa-associated lymphoid tissue lymphoma translocation protein 1-like	7.93
28	991289	1002123	tight junction protein 3	7.93
28	1036061	1041471	cactin, spliceosome C complex subunit	7.93
28	1041856	1045661	thromboxane A2 receptor	7.93
28	1047302	1049378	GIPC PDZ domain containing family, member 3	7.93
28	1049765	1054525	high mobility group 20B	7.93
28	1086890	1089949	Deoxyhypusine hydroxylase	7.93
28	1102700	1127154	unc-13 homolog A (C. elegans)	7.93
28	1143774	1145084	phosphatidic acid phosphatase type 2C	7.93
28	1171971	1179236	amino-terminal enhancer of split	7.93
28	1195931	1202586	guanine nucleotide-binding protein subunit alpha-11	7.93
28	1218202	1225645	nicalin precursor	7.93
28	1233711	1264738	CUGBP, Elav-like family member 5	7.93
28	1266179	1271107	Hydroxysteroid 11-beta-dehydrogenase 1-like protein	7.93
28	1303285	1407097	MPN domain containing	7.93
28	1414197	1415393	Zinc finger and BTB domain-containing protein 7A	7.93
28	1420886	1434153	protein inhibitor of activated STAT, 4	7.93
28	1438607	1446530	elongation factor 2	7.93
28	1447842	1453358	death-associated protein kinase 3	7.93
28	1455327	1459014	nicotinamide riboside kinase 2	7.93
28	1463358	1478145	ataxia, cerebellar, Cayman type	7.93
28	1507549	1509672	neurturin	7.93
28	1442292	1442361	Small nucleolar RNA SNORD37	7.93

Supplementary Table 8. List of genes for selective sweeps detected with hapFLK with 0.05% threshold in white layers.

Chr	Start	End	Description	hapFLK
1	8177104	8318907	Semaphorin-3D	0.46
1	8522175	8850302	semaphorin-3A	0.46
1	32863934	32927887	ubiquitin carboxyl-terminal hydrolase 15 isoform 1	0.46
1	32942077	33016611	protein MON2 homolog	0.46
1	33019230	33019319	gga-let-7i	0.46
1	156898969	157114305	kelch-like family member 1	0.46
1	161506641	161608657	tudor domain-containing protein 3	0.45
1	161696600	161932705	diaphanous homolog 3 (Drosophila)	0.45
2	73536943	73537003	Small nucleolar RNA R11/Z151	0.46
2	95477061	95554752	cadherin-7 precursor	0.46
3	60548727	60598139	clavesin 2	0.46
3	60674294	60677648	Fatty acid-binding protein, brain	0.46
3	60687661	60743857	protein kinase (cAMP-dependent, catalytic) inhibitor beta	0.46
3	60764156	60778623	serine incorporator 1 precursor	0.46
3	60779761	60800976	Heat shock factor protein 2	0.46
3	61222961	61230461	gap junction alpha-1 protein	0.46
4	52775238	52982663	protein sprouty homolog 1	0.46
4	52983308	52995163	nudix (nucleoside diphosphate linked moiety X)-type motif 6	0.46
4	52994729	53020450	Fibroblast growth factor 2	0.46
4	53097087	53101842	interleukin 21 precursor	0.46
4	53134199	53137244	interleukin 2 precursor	0.46
4	53155161	53264068	KIAA1109	0.46
4	53335689	53350682	Bardet-Biedl syndrome 7	0.46
4	53350909	53356114	cyclin A2	0.46
6	14411231	14417212	voltage-dependent anion-selective channel protein 2	4.95
6	14522141	14541652	Dual specificity phosphatase DUPD1	4.95
6	14667844	14924979	adenosine kinase	4.95
6	14941845	14956153	AP-3 complex subunit mu-1	4.95
6	14961440	14990291	Vinculin	4.95
6	15060198	15068452	urokinase-type plasminogen activator preproprotein	4.95
6	24664740	24940362	sortilin-related VPS10 domain containing receptor 1	4.95
6	25264735	25352280	gamma-adducin	4.92
6	25366993	25409141	max-interacting protein 1	4.92
6	25517714	25527369	Dual specificity protein phosphatase	4.92
6	25533210	25555268	structural maintenance of chromosomes protein 3	4.92
6	25683192	25695299	programmed cell death protein 4	4.92
9	11379753	11384969	Zic family member 4	0.46
9	11938121	11940148	phosphatidylinositol glycan anchor biosynthesis, class Z	0.46
9	11940276	11956868	melanotransferrin precursor	0.46
9	11959073	12089444	discs, large homolog 1 (Drosophila)	0.46
9	12115163	12127577	D-beta-hydroxybutyrate dehydrogenase, mitochondrial precursor	0.46
9	12153804	12156739	apolipoprotein D precursor	0.46
9	12157987	12170689	protein phosphatase inhibitor 2	0.46
9	12171478	12231009	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein	0.46
9	12296460	12308388	large subunit GTPase 1 homolog	0.46

9	11878868	11880012	Type-1 angiotensin II receptor	0.46
13	1767729	1785069	endothelial cell surface expressed chemotaxis and apoptosis regulator	5.33
13	1788284	1799938	dnaJ homolog subfamily C member 18	5.33
13	1806725	1808111	marginal zone B and B1 cell-specific protein	5.33
13	1813947	1819364	Polyadenylate-binding protein-interacting protein 2	5.33
13	1825981	1851149	matrin-3	5.33
13	2010626	2012530	leucine rich repeat transmembrane neuronal 2	5.33
13	2134785	2156387	stress-70 protein, mitochondrial precursor	5.33
13	2567435	2611832	fibroblast growth factor 18 precursor	5.33
13	2641231	2651587	nucleophosmin	5.33
13	2681554	2828621	RAN binding protein 17	5.33
13	2146299	2146368	Small nucleolar RNA SNORD63	5.33

Supplementary Table 9. List of genes for selective sweeps detected with hapFLK with 0.05% threshold in brown layers.

Chr	Start	End	Description	hapFLK
1	50719186	50733066	DMC1 dosage suppressor of mck1 homolog, meiosis-specific homologous	6.09
1	50753266	50759412	KDEL receptor 3	6.09
1	50770322	50773091	Inward-rectifying potassium channel cKir2.3; Uncharacterized protein	6.09
1	50788491	50804577	casein kinase I isoform epsilon	6.09
1	50832960	50835917	Transcription factor MafF	6.09
1	50836439	50854357	85 kDa calcium-independent phospholipase A2	6.09
1	50856306	50864314	BAI1-associated protein 2-like 2	6.09
1	50865712	50873377	Monocarboxylate transporter 3	6.09
1	50911765	50921215	Transcription factor SOX-10	6.09
1	50924002	50927570	DNA-directed RNA polymerases I, II, and III subunit RPABC2	6.09
1	50931992	50950143	MICAL-like protein 1	6.09
1	50953763	50963376	Eukaryotic translation initiation factor 3 subunit L	6.09
1	50979163	50980997	galanin receptor type 3	6.09
1	51010961	51014093	Beta-galactoside-binding lectin	6.09
1	51026210	51029611	pyridoxal (pyridoxine, vitamin B6) phosphatase	6.09
1	51050593	51062352	lectin, galactoside-binding, soluble, 2	6.09
1	51065196	51067106	CDC42 effector protein (Rho GTPase binding) 1	6.09
1	51083236	51095519	caspase recruitment domain family, member 10	6.09
1	51099652	51110327	MFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	6.09
1	51229151	51248424	cytohesin-4	6.09
1	51257450	51266551	ras-related C3 botulinum toxin substrate 2	6.09
1	51269724	51274129	somatostatin receptor type 3	6.09
1	51336790	51344019	potassium channel tetramerisation domain containing 17	6.09
1	51349989	51355027	Sulfurtransferase	6.09
1	51355166	51362344	thiosulfate sulfurtransferase	6.09
1	51419978	51430209	neutrophil cytosol factor 4	6.09
1	51455040	51463393	Parvalbumin, muscle	6.09
1	51573280	51580672	thioredoxin, mitochondrial	6.09
1	51598550	51666512	myosin-9	6.09
1	50983576	50984295	noggin 4 precursor	6.09
1	50991976	50992548	Histone H5	6.09
1	51511551	51511715	TUC338	6.09
2	113502155	113517713	tocopherol (alpha) transfer protein	0.75

2	113543309	113551391	YTH domain family protein 3	0.75
2	114472659	114503139	Armadillo repeat-containing protein 1	0.75
2	113888475	113888659	TUC338	0.75
3	6450555	7094776	neurexin-1-alpha isoform 1 precursor	0.75
7	23919267	24182349	Contactin-associated protein-like 5	0.76
10	6234433	6311306	myosin IE	0.76
10	6313607	6321500	G2/mitotic-specific cyclin-B2	0.76
10	6322555	6364720	E3 ubiquitin-protein ligase Arkadia	0.76
10	6377333	6395463	SAFB-like, transcription modulator	0.76
10	6434758	6619897	disintegrin and metalloproteinase domain-containing protein 10 precursor	0.76
10	6480269	6521283	lipase, hepatic	0.76
10	6572777	6596733	aquaporin 9	0.76
10	6604818	6661741	Retinal dehydrogenase 2	0.76
10	6816944	6857906	cingulin-like 1	0.76
10	6903270	7058903	transcription factor 12	0.76
10	7158297	7164937	meiosis-specific nuclear structural 1	0.76
10	7159553	7183472	testis expressed 9	0.76
10	7249251	7256715	regulatory factor X, 7	0.76
10	7266832	7316503	E3 ubiquitin-protein ligase	0.76
11	12288497	12432361	cadherin-8	0.76
11	13255162	13323250	ADAM metallopeptidase with thrombospondin type 1 motif, 18	0.76
11	12638991	12639143	TUC338	0.76
13	10000421	10264193	transcription factor COE1	0.75
13	10554868	10603091	clathrin interactor 1	0.75
13	10618933	10623222	LSM11, U7 small nuclear RNA associated	0.75
13	10624187	10630514	probable tRNA(His) guanylyltransferase	0.75
13	10642122	10669352	disintegrin and metalloproteinase domain-containing protein 19 precursor	0.75
13	10668257	10680834	NIPA-like domain containing 4	0.75
13	10737034	10764716	IL2-inducible T-cell kinase	0.75
13	10778915	10785137	hepatitis A virus cellular receptor 1 precursor	0.75
20	10382093	10392964	targeting protein for Xklp2	0.76
20	10395495	10403016	myosin light chain kinase 2, skeletal/cardiac muscle	0.76
20	10405912	10410873	interferon regulatory factor 10	0.76
20	10412313	10419874	dual specificity phosphatase 15	0.76
20	10420572	10431015	tubulin tyrosine ligase-like family, member 9	0.76
20	10431992	10433839	p53 and DNA-damage regulated 1	0.76
20	10487518	10491477	GDP-fucose protein O-fucosyltransferase 1 precursor	0.76
20	10492514	10502873	kinesin-like protein KIF3B	0.76
20	10505808	10517025	additional sex combs like 1 (Drosophila)	0.76
20	10595952	10609143	DNA (cytosine-5-)-methyltransferase 3 beta	0.76
20	10610165	10617328	Microtubule-associated protein RP/EB family member 1	0.76
20	10633103	10644913	Ovocalyxin-36 precursor	0.76
20	10643158	10647278	protein TENP	0.76
20	10649024	10653537	BPI fold containing family B, member 6	0.76
20	10663664	10668879	BPI fold containing family B, member 4	0.76
20	10700368	10707694	KIAA1755	0.76
20	10714657	10723053	protein-glutamine gamma-glutamyltransferase 2	0.76
20	10726664	10747756	regulation of nuclear pre-mRNA domain-containing protein 1B	0.76
20	10772139	10813719	catenin, beta like 1	0.76

20	10849414	10853705	deoxynucleotidyltransferase, terminal, interacting protein 1	0.76
20	10856323	10858120	Troponin C, skeletal muscle	0.76
20	10871289	10873861	neuralized homolog 2 (Drosophila)	0.76
20	10874294	10878191	lysosomal protective protein precursor	0.76
20	10878671	10881926	phospholipid transfer protein precursor	0.76
20	10887208	10895373	PDX1 C-terminal inhibiting factor 1	0.76
20	10917368	10921453	matrix metalloproteinase-9 precursor	0.76
20	10922030	10944646	solute carrier family 12 (potassium/chloride transporter), member 5	0.76
20	10966821	10970503	tumor necrosis factor receptor superfamily member 5 precursor	0.76
20	11071603	11076564	solute carrier family 35 member C2	0.76
26	4133855	4213476	ankyrin repeat and sterile alpha motif domain containing 1A	6.05
26	4214176	4217459	transcription initiation factor TFIID subunit 11	6.05
26	4248039	4252009	U1 small nuclear ribonucleoprotein C	6.05
26	4302444	4306682	protein kinase C and casein kinase substrate in neurons 1	6.05
26	4380152	4414660	glutamate receptor, metabotropic 4	6.05
26	4496441	4502013	Green-sensitive opsin	6.05
26	4515351	4518565	Motilin	6.05
26	4540474	4553563	inositol hexakisphosphate kinase 3	6.05
26	4560707	4598750	inositol 1,4,5-trisphosphate receptor, type 3	6.05
26	4657007	4660254	O-acetyl-ADP-ribose deacetylase C6orf130 homolog	6.05
26	4660732	4674975	nuclear transcription factor Y subunit alpha	6.05
26	4678907	4684706	Triggering receptor expressed on myeloid cells; Uncharacterized protein	6.05
26	4686588	4691130	triggering receptor expressed on myeloid cells 2 precursor	6.05
26	4693285	4699278	triggering receptor expressed on myeloid cells precursor	6.05
26	4891507	4916817	forkhead box P4	6.05
26	4960037	4974337	transcription factor EB	6.05
26	4976148	4979037	progastricsin (pepsinogen C)	6.05
26	4980962	4984806	gastricsin precursor	6.05
26	4989811	4997320	fibroblast growth factor receptor substrate 3	6.05
26	5001119	5006673	prickle homolog 4 (Drosophila)	6.05
26	5028663	5038433	Ubiquitin carboxyl-terminal hydrolase	6.05
26	5056220	5060765	Mediator of RNA polymerase II transcription subunit 20	6.05
26	5060817	5063516	bystin-like	6.05
26	5065132	5077551	G1/S-specific cyclin-D3	6.05
26	5103140	5110535	Transcription initiation factor TFIID subunit 8	6.05
26	5112283	5114932	primary cilia formation	6.05
26	5122698	5126812	acidic chitinase precursor	6.05
27	3080955	3092426	histone acetyltransferase MYST2	5.65
27	3121453	3124525	solute carrier family 35 member B1	5.65
27	3204183	3219385	tumor necrosis factor receptor superfamily member 16 precursor	5.65
27	3252521	3267354	membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3)	5.65
27	3315353	3321406	homeobox protein MOX-1	5.65
27	3335561	3350183	ets variant 4	5.65
27	3352213	3364975	DEAH (Asp-Glu-Ala-His) box polypeptide 8	5.65
27	3366784	3370919	Prohibitin	5.65
27	3413412	3417057	phosphoethanolamine/phosphocholine phosphatase	5.65
27	3423443	3425719	guanine nucleotide binding protein, gamma transducing activity polypeptide 2	5.65
27	3433057	3459259	Insulin-like growth factor 2 mRNA-binding protein 1	5.65
27	3468107	3475761	gastric inhibitory polypeptide precursor	5.65

27	3495909	3506537	calcium binding and coiled-coil domain 2	5.65
27	3586132	3589844	Hoxb-7	5.65
27	3598354	3600970	homeobox B6	5.65
27	3604171	3606516	Homeobox protein Hox-B5	5.65
27	3621645	3626317	homeobox protein Hox-B4	5.65
27	3643412	3649970	homeobox protein Hox-B3	5.65
27	3652538	3655356	homeobox B2	5.65
27	3662963	3664600	homeobox B1	5.65
27	3711055	3815879	src kinase associated phosphoprotein 1	5.65
27	3842373	3850380	chromobox protein homolog 1	5.65
27	3855203	3861688	Nuclear factor erythroid 2-related factor 1	5.65
27	3865615	3868783	CDK5 regulatory subunit associated protein 3	5.65
27	3896638	3899664	leucine rich repeat containing 46	5.65
27	3902862	3907297	oxysterol binding protein-like 7	5.65
27	3971094	4005910	myeloid/lymphoid or mixed-lineage leukemia ; translocated to, 6	5.65
27	4006813	4007489	CDGSH iron sulfur domain 3	5.65
27	4008146	4011942	polycomb group ring finger 2	5.65
27	4014223	4017030	Proteasome subunit beta type	5.65
27	3869421	3870521	proline rich 15-like	5.65
27	3155867	3156043	TUC338	5.65
27	4685681	4690379	Small nuclear ribonucleoprotein-associated protein B	0.76
27	4698639	4700783	kelch-like family member 11	0.76
27	4701712	4720355	ATP-citrate synthase	0.76
27	4739049	4758336	dnaJ homolog subfamily C member 7	0.76
27	4758782	4759794	NF-kappa-B inhibitor-interacting Ras-like protein 2	0.76
27	4822592	4826644	DEXH (Asp-Glu-X-His) box polypeptide 58	0.76
27	4827918	4832197	histone acetyltransferase KAT2A	0.76
27	4837933	4841685	Ras-related protein Rab-5C	0.76
27	4847654	4856867	potassium voltage-gated channel, subfamily H (eag-related), member 4	0.76
27	4857256	4858313	hypocretin (orexin) neuropeptide precursor	0.76
27	4871019	4873562	GH3 domain containing	0.76
27	4875164	4886606	signal transducer and activator of transcription 5B	0.76
27	4896267	4907552	Signal transducer and activator of transcription 3	0.76
27	4913996	4926852	polymerase I and transcript release factor	0.76
27	4929945	4958104	V-type proton ATPase 116 kDa subunit a isoform 1	0.76
27	4962219	4963268	17-beta-hydroxysteroid dehydrogenase	0.76
27	4963812	4965610	CoA synthase	0.76
27	4967021	4970406	max-like protein X	0.76
27	4970565	4973603	PSMC3 interacting protein	0.76
27	4990263	4995457	pleckstrin homology domain containing, family H member 3	0.76
27	5003802	5012710	contactin associated protein 1	0.76
27	5030186	5032034	receptor activity-modifying protein 2 precursor	0.76
27	5039150	5051893	WNK lysine deficient protein kinase 4	0.76
27	5055739	5060519	beclin-1	0.76
27	5060694	5067398	Proteasome activator complex subunit 3	0.76
27	5095037	5098091	RUN domain containing 1	0.76
27	5099248	5101501	60S ribosomal protein L27	0.76
27	5102844	5105307	interferon-induced protein 35	0.76
27	5126612	5140086	rho-related GTP-binding protein RhoN	0.76

27	5150701	5171059	breast cancer 1, early onset	0.76
27	5171677	5189435	neighbor of BRCA1 gene 1	0.76
27	5192609	5198967	membrane protein, palmitoylated 2 (MAGUK p55 subfamily member 2)	0.76
27	5000132	5001181	chemokine (C-C motif) receptor 10	0.76
27	5173809	5173871	Neighbour of BRCA1 gene 2 converved region	0.76

Supplementary Table 10. Lists of pathways and gene ontologies under selection with hapFLK with 0.05% threshold in all layers.

Description	# Genes anotated	Genes of pathways (%)*	P-Value
protein binding	2	2.0	0.000
apical junction assembly	2	28.6	0.000
epithelial cell-cell adhesion	2	28.6	0.000
cadherin binding	2	14.3	0.001
adherens junction	2	10.5	0.001
cell-cell adherens junction	2	9.1	0.002
mitochondrial nucleoid	2	7.4	0.002
integrin activation	1	50.0	0.005
intracellular organelle	1	50.0	0.005
L-ascorbic acid transport	1	50.0	0.005
L-ascorbic acid transporter activity	1	50.0	0.005
mediator complex binding	1	50.0	0.005
meiotic cohesin complex	1	50.0	0.005
negative regulation of integrin-mediated signaling pathway	1	50.0	0.005
positive regulation of RNA splicing	1	50.0	0.005
regulation of endodeoxyribonuclease activity	1	50.0	0.005
regulation of relaxation of cardiac muscle	1	50.0	0.005
sodium-dependent L-ascorbate transmembrane transporter activity	1	50.0	0.005
transepithelial L-ascorbic acid transport	1	50.0	0.005
beta-catenin binding	2	4.7	0.006
nucleosome assembly	2	4.3	0.007
bradykinin catabolic process	1	33.3	0.008
calcium-dependent cell-cell adhesion	1	33.3	0.008
cellular response to indole-3-methanol	1	33.3	0.008
dehydroascorbic acid transporter activity	1	33.3	0.008
endoplasmic reticulum chaperone complex	1	33.3	0.008
negative regulation of DNA endoreduplication	1	33.3	0.008
nuclear meiotic cohesin complex	1	33.3	0.008
regulation of B cell proliferation	1	33.3	0.008
regulation of receptor activity	1	33.3	0.008
translation release factor activity, codon specific	1	33.3	0.008
type 1 fibroblast growth factor receptor binding	1	33.3	0.008
type 2 fibroblast growth factor receptor binding	1	33.3	0.008
zonula adherens	1	33.3	0.008
adherens junction assembly	1	25.0	0.011
AMP biosynthetic process	1	25.0	0.011
dehydroascorbic acid transport	1	25.0	0.011
fibrinolysis	1	25.0	0.011
negative regulation of cell motility	1	25.0	0.011
negative regulation of centrosome duplication	1	25.0	0.011

negative regulation of protein kinase activity	1	25.0	0.011
regulation of smooth muscle cell migration	1	25.0	0.011
smooth muscle cell migration	1	25.0	0.011
translation repressor activity	1	25.0	0.011
translational termination	1	25.0	0.011
voltage-gated anion channel activity	1	25.0	0.011
alpha-catenin binding	1	20.0	0.013
dystroglycan binding	1	20.0	0.013
labyrinthine layer development	1	20.0	0.013
MOZ/MORF histone acetyltransferase complex	1	20.0	0.013
negative regulation of neuroblast proliferation	1	20.0	0.013
negative regulation of translational initiation	1	20.0	0.013
neuron fate specification	1	20.0	0.013
purine ribonucleoside salvage	1	20.0	0.013
regulation of centriole replication	1	20.0	0.013
regulation of respiratory gaseous exchange by neurological system	1	20.0	0.013
Tat protein binding	1	20.0	0.013
calmodulin-dependent protein kinase activity	1	16.7	0.016
catenin complex	1	16.7	0.016
establishment or maintenance of cell polarity	1	16.7	0.016
lateral element	1	16.7	0.016
mitotic spindle organization	1	16.7	0.016
ribosomal small subunit binding	1	16.7	0.016
spindle pole centrosome	1	16.7	0.016
cell-cell junction	2	2.7	0.017
gamma-catenin binding	1	14.3	0.019
regulation of anion transport	1	14.3	0.019
phosphotransferase activity, alcohol group as acceptor	2	2.4	0.021
MAP kinase tyrosine/serine/threonine phosphatase activity	1	12.5	0.021
O-methyltransferase activity	1	12.5	0.021
positive regulation of extrinsic apoptotic signaling pathway	1	12.5	0.021
protein kinase inhibitor activity	1	12.5	0.021
regulation of DNA replication	1	12.5	0.021
vinculin binding	1	12.5	0.021
acetyltransferase activity	1	11.1	0.024
costamere	1	11.1	0.024
regulation of cell adhesion mediated by integrin	1	11.1	0.024
regulation of cell proliferation	2	2.2	0.024
structural molecule activity	2	2.2	0.025
fascia adherens	1	10.0	0.027
MAP kinase activity	1	10.0	0.027
dynein binding	1	9.1	0.029
meiosis	1	9.1	0.029
actin cytoskeleton	2	2.0	0.030
nucleoplasm	2	1.9	0.032
basal plasma membrane	1	8.3	0.032
ceramide biosynthetic process	1	8.3	0.032
protein destabilization	1	8.3	0.032
SCF-dependent proteasomal ubiquitin-dependent protein catabolic	1	8.3	0.032

anterograde synaptic vesicle transport	1	7.7	0.035
morphogenesis of an epithelium	1	7.7	0.035
respiratory gaseous exchange	1	7.7	0.035
nucleus	2	2.1	0.037
anterograde axon cargo transport	1	7.1	0.037
glycogen metabolic process	1	7.1	0.037
protein export from nucleus	1	7.1	0.037
protein localization to cell surface	1	7.1	0.037
anion transport	1	6.7	0.040
cell fate specification	1	6.7	0.040
histone acetylation	1	6.7	0.040
positive regulation of proteolysis	1	6.7	0.040
positive regulation of smoothened signaling pathway	1	6.7	0.040
protein dimerization activity	2	1.7	0.042
aminopeptidase activity	1	6.3	0.042
cell aging	1	6.3	0.042
RNA metabolic process	1	6.3	0.042
sodium ion transmembrane transport	1	6.3	0.042
negative regulation of intrinsic apoptotic signaling pathway	1	5.9	0.045
protein methylation	1	5.9	0.045
intercalated disc	1	5.6	0.048
negative regulation of extrinsic apoptotic signaling pathway	1	5.6	0.048

*Percentage of the genes of the pathway which were among the annotated genes.

Supplementary Table 11. Lists of pathways and gene ontologies under selection with hapFLK with 0.05% threshold in white layers.

Description	**# Genes anotated**	**Genes of pathways (%)***	**P-Value**
protein binding	2	2.0	0.000
apical junction assembly	2	28.6	0.000
epithelial cell-cell adhesion	2	28.6	0.000
cadherin binding	2	14.3	0.001
adherens junction	2	10.5	0.001
cell-cell adherens junction	2	9.1	0.002
mitochondrial nucleoid	2	7.4	0.002
integrin activation	1	50.0	0.005
intracellular organelle	1	50.0	0.005
L-ascorbic acid transport	1	50.0	0.005
L-ascorbic acid transporter activity	1	50.0	0.005
mediator complex binding	1	50.0	0.005
meiotic cohesin complex	1	50.0	0.005
negative regulation of integrin-mediated signaling pathway	1	50.0	0.005
positive regulation of RNA splicing	1	50.0	0.005
regulation of endodeoxyribonuclease activity	1	50.0	0.005
regulation of relaxation of cardiac muscle	1	50.0	0.005
sodium-dependent L-ascorbate transmembrane transporter activity	1	50.0	0.005
transepithelial L-ascorbic acid transport	1	50.0	0.005
beta-catenin binding	2	4.7	0.006
nucleosome assembly	2	4.3	0.007

bradykinin catabolic process	1	33.3	0.008
calcium-dependent cell-cell adhesion	1	33.3	0.008
cellular response to indole-3-methanol	1	33.3	0.008
dehydroascorbic acid transporter activity	1	33.3	0.008
endoplasmic reticulum chaperone complex	1	33.3	0.008
negative regulation of DNA endoreduplication	1	33.3	0.008
nuclear meiotic cohesin complex	1	33.3	0.008
regulation of B cell proliferation	1	33.3	0.008
regulation of receptor activity	1	33.3	0.008
translation release factor activity, codon specific	1	33.3	0.008
type 1 fibroblast growth factor receptor binding	1	33.3	0.008
type 2 fibroblast growth factor receptor binding	1	33.3	0.008
zonula adherens	1	33.3	0.008
adherens junction assembly	1	25.0	0.011
AMP biosynthetic process	1	25.0	0.011
dehydroascorbic acid transport	1	25.0	0.011
fibrinolysis	1	25.0	0.011
negative regulation of cell motility	1	25.0	0.011
negative regulation of centrosome duplication	1	25.0	0.011
negative regulation of protein kinase activity	1	25.0	0.011
regulation of smooth muscle cell migration	1	25.0	0.011
smooth muscle cell migration	1	25.0	0.011
translation repressor activity	1	25.0	0.011
translational termination	1	25.0	0.011
voltage-gated anion channel activity	1	25.0	0.011
alpha-catenin binding	1	20.0	0.013
dystroglycan binding	1	20.0	0.013
labyrinthine layer development	1	20.0	0.013
MOZ/MORF histone acetyltransferase complex	1	20.0	0.013
negative regulation of neuroblast proliferation	1	20.0	0.013
negative regulation of translational initiation	1	20.0	0.013
neuron fate specification	1	20.0	0.013
purine ribonucleoside salvage	1	20.0	0.013
regulation of centriole replication	1	20.0	0.013
regulation of respiratory gaseous exchange	1	20.0	0.013
Tat protein binding	1	20.0	0.013
calmodulin-dependent protein kinase activity	1	16.7	0.016
catenin complex	1	16.7	0.016
establishment or maintenance of cell polarity	1	16.7	0.016
lateral element	1	16.7	0.016
mitotic spindle organization	1	16.7	0.016
ribosomal small subunit binding	1	16.7	0.016
spindle pole centrosome	1	16.7	0.016
cell-cell junction	2	2.7	0.017
gamma-catenin binding	1	14.3	0.019
regulation of anion transport	1	14.3	0.019
phosphotransferase activity, alcohol group as acceptor	2	2.4	0.021
MAP kinase tyrosine/serine/threonine phosphatase activity	1	12.5	0.021
O-methyltransferase activity	1	12.5	0.021

positive regulation of extrinsic apoptotic signaling pathway	1	12.5	0.021
protein kinase inhibitor activity	1	12.5	0.021
regulation of DNA replication	1	12.5	0.021
vinculin binding	1	12.5	0.021
acetyltransferase activity	1	11.1	0.024
costamere	1	11.1	0.024
regulation of cell adhesion mediated by integrin	1	11.1	0.024
regulation of cell proliferation	2	2.2	0.024
structural molecule activity	2	2.2	0.025
fascia adherens	1	10.0	0.027
MAP kinase activity	1	10.0	0.027
dynein binding	1	9.1	0.029
meiosis	1	9.1	0.029
actin cytoskeleton	2	2.0	0.030
nucleoplasm	2	1.9	0.032
basal plasma membrane	1	8.3	0.032
ceramide biosynthetic process	1	8.3	0.032
protein destabilization	1	8.3	0.032
SCF-dependent proteasomal ubiquitin-dependent protein catabolic	1	8.3	0.032
anterograde synaptic vesicle transport	1	7.7	0.035
morphogenesis of an epithelium	1	7.7	0.035
respiratory gaseous exchange	1	7.7	0.035
nucleus	2	2.1	0.037
anterograde axon cargo transport	1	7.1	0.037
glycogen metabolic process	1	7.1	0.037
protein export from nucleus	1	7.1	0.037
protein localization to cell surface	1	7.1	0.037
anion transport	1	6.7	0.040
cell fate specification	1	6.7	0.040
histone acetylation	1	6.7	0.040
positive regulation of proteolysis	1	6.7	0.040
positive regulation of smoothened signaling pathway	1	6.7	0.040
protein dimerization activity	2	1.7	0.042
aminopeptidase activity	1	6.3	0.042
cell aging	1	6.3	0.042
RNA metabolic process	1	6.3	0.042
sodium ion transmembrane transport	1	6.3	0.042
negative regulation of intrinsic apoptotic signaling pathway	1	5.9	0.045
protein methylation	1	5.9	0.045
intercalated disc	1	5.6	0.048
negative regulation of extrinsic apoptotic signaling pathway	1	5.6	0.048

*Percentage of the genes of the pathway which were among the annotated genes.

Supplementary Table 12. Lists of pathways and gene ontologies under selection with hapFLK with 0.05% threshold in brown layers.

Description	# Genes anotated	Genes of pathways (%)*	P-Value
RNA polymerase II transcription factor	2	18.2	0.005
negative regulation of MAPK cascade	2	16.7	0.005
immunological synapse	2	14.3	0.007
inward rectifier potassium channel activity	2	12.5	0.010
RNA polymerase II distal enhancer sequence-specific DNA binding	2	12.5	0.010
phosphatidylinositol metabolic process	2	11.8	0.011
cholesterol binding	2	10.0	0.015
ruffle membrane	2	9.5	0.016
phosphatidylinositol binding	3	4.8	0.021
sequence-specific DNA binding RNA polymerase II transcription	2	7.4	0.026
skeletal muscle cell differentiation	2	6.7	0.032
negative regulation of Wnt receptor signaling pathway	2	6.3	0.036
Rab GTPase binding	2	5.7	0.043

*Percentage of the genes of the pathway which were among the annotated genes.

Reference

1. Kaplan NL, Hudson RR, Langley CH (1989) The “hitchhiking effect” revisited. Genetics 123: 887–899.

2. Nielsen R (2005) Molecular Signatures of Natural Selection. Annu Rev Genet 39: 197–218. doi:10.1146/annurev.genet.39.073003.112420.

3. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832–837. doi:10.1038/nature01140.

4. Qanbari S, Simianer H (n.d.) Mapping signatures of positive selection in the genome of livestock. Livest Sci. Available: http://www.sciencedirect.com/science/article/pii/S187114131400239X. Accessed 18 June 2014.

5. Kim Y, Nielsen R (2004) Linkage Disequilibrium as a Signature of Selective Sweeps. Genetics 167: 1513–1524. doi:10.1534/genetics.103.025387.

6. Lewontin RC, Krakauer J (1973) Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms. Genetics 74: 175–195.

7. Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13: 969–980.

8. Innan H, Kim Y (2008) Detecting Local Adaptation Using the Joint Sampling of Polymorphism Data in the Parental and Derived Populations. Genetics 179: 1713–1720. doi:10.1534/genetics.108.086835.

9. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, et al. (2010) Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science 329: 75–78. doi:10.1126/science.1190371.

10. Wright S (1949) The Genetical Structure of Populations. Ann Eugen 15: 323–354. doi:10.1111/j.1469-1809.1949.tb02451.x.

11. Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. Nat Rev Genet 11: 459–463. doi:10.1038/nrg2813.

12. Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, et al. (2010) Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. Genetics 186: 241–262. doi:10.1534/genetics.110.117275.

13. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, et al. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449: 913–918. doi:10.1038/nature06250.

14. Browning SR, Weir BS (2010) Population Structure With Localized Haplotype Clusters. Genetics 185: 1337–1344. doi:10.1534/genetics.110.116681.

15. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193: 929–941. doi:10.1534/genetics.112.147231.

16. Brown WRA, Hubbard SJ, Tickle C, Wilson SA (2003) The chicken as a model for large-scale analysis of vertebrate gene function. Nat Rev Genet 4: 87–98. doi:10.1038/nrg998.

17. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, et al. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464: 587–591. doi:10.1038/nature08832.

18. Johansson AM, Pettersson ME, Siegel PB, Carlborg Ö (2010) Genome-Wide Effects of Long-Term Divergent Selection. PLoS Genet 6: e1001188. doi:10.1371/journal.pgen.1001188.

19. Qanbari S, Strom TM, Haberer G, Weigend S, Gheyas AA, et al. (2012) A High Resolution Genome-Wide Scan for Significant Selective Sweeps: An Application to Pooled Sequence Data in Laying Chickens. PLoS ONE 7: e49525. doi:10.1371/journal.pone.0049525.

20. Gholami M, Erbe M, Gärke C, Preisinger R, Weigend A, et al. (2014) Population Genomic Analyses Based on 1 Million SNPs in Commercial Egg Layers. PLoS ONE 9: e94509. doi:10.1371/journal.pone.0094509.

21. Crawford R (1990) poultry breeding and genetics. Elsevier science.

22. Burt DW (2005) Chicken genome: Current status and future opportunities. Genome Res 15: 1692–1698. doi:10.1101/gr.4141805.

23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81: 559–575.

24. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. Genetics 105: 767–779.

25. Zhou H, Alexander D, Lange K (2011) A quasi-Newton acceleration for high-dimensional optimization algorithms. Stat Comput 21: 261–273. doi:10.1007/s11222-009-9166-3.

26. Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78: 629–644. doi:10.1086/502802.

27. Clarke BR, McKinnon PL, Riley G (2012) A fast robust method for fitting gamma distributions. Stat Pap 53: 1001–1014. doi:10.1007/s00362-011-0404-3.

28. Durinck S (n.d.) biomaRt: Interface to BioMart databases (e.g. Ensembl, COSMIC ,Wormbase and Gramene). R package version 260.

29. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, et al. (2012) Ensembl 2013. Nucleic Acids Res 41: D48–D55. doi:10.1093/nar/gks1236.

30. Hu Z-L, Park CA, Wu X-L, Reecy JM (2012) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res 41: D871–D879. doi:10.1093/nar/gks1150.

31. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40: D109–114. doi:10.1093/nar/gkr988.

32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29. doi:10.1038/75556.

33. Fisher RA (1922) On the Interpretation of $\chi 2$ from Contingency Tables, and the Calculation of P. J R Stat Soc 85: 87–94. doi:10.2307/2340521.

34. Xu P, Denbow CJ, Meiri N, Denbow DM (2012) Fasting of 3-day-old chicks leads to changes in histone H3 methylation status. Physiol Behav 105: 276–282. doi:10.1016/j.physbeh.2011.06.023.

35. Bai Y, Sun G, Kang X, Han R, Tian Y, et al. (2012) Polymorphisms of the pro-opiomelanocortin and agouti-related protein genes and their association with chicken production traits. Mol Biol Rep 39: 7533–7539. doi:10.1007/s11033-012-1587-y.

36. Kim S, Miska KB, McElroy AP, Jenkins MC, Fetterer RH, et al. (2009) Molecular cloning and functional characterization of avian interleukin-19. Mol Immunol 47: 476–484. doi:10.1016/j.molimm.2009.08.027.

37. Lei M, Peng X, Zhou M, Luo C, Nie Q, et al. (2008) Polymorphisms of the IGF1R gene and their genetic effects on chicken early growth and carcass traits. BMC Genet 9: 70. doi:10.1186/1471-2156-9-70.

38. Zhao XH, Wang JY, Zhang GX, Wei Y, Gu YP, et al. (2012) Single nucleotide polymorphism in the STAT5b gene is associated with body weight and reproductive traits of the Jinghai Yellow chicken. Mol Biol Rep 39: 4177–4183. doi:10.1007/s11033-011-1202-7.

39. Gautron J, Réhault-Godbert S, Pascal G, Nys Y, Hincke MT (2011) Ovocalyxin-36 and other LBP/BPI/PLUNC-like proteins as molecular actors of the mechanisms of the avian egg natural defences. Biochem Soc Trans 39: 971–976. doi:10.1042/BST0390971.

40. Gunnarsson U, Kerje S, Bed'hom B, Sahlqvist A-S, Ekwall O, et al. (2011) The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. Pigment Cell Melanoma Res 24: 268–274. doi:10.1111/j.1755-148X.2011.00825.x.

41. Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. Mol Ecol Resour 11: 184–194. doi:10.1111/j.1755-0998.2011.02987.x.

42. Akey JM (2009) Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res 19: 711–722. doi:10.1101/gr.086652.108.

43. Nuzhdin SV, Turner TL (2013) Promises and limitations of hitchhiking mapping. Curr Opin Genet Dev 23: 694–699. doi:10.1016/j.gde.2013.10.002.

44. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23: 23–35. doi:10.1017/S0016672300014634.

45. Pritchard JK, Pickrell JK, Coop G (2010) The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. Curr Biol 20: R208–R215. doi:10.1016/j.cub.2009.11.055.

46. Nielsen R, Slatkin M (2013) An introduction to population genetics: theory and applications. Sunderland, Mass.: Sinauer Associates.

47. Qanbari S, Hansen M, Weigend S, Preisinger R, Simianer H (2010) Linkage disequilibrium reveals different demographic history in egg laying chickens. BMC Genet 11: 103. doi:10.1186/1471-2156-11-103.

48. Muir WM, Aggrey SE (2003) Poultry Genetics, Breeding and Biotechnology. CABI. 724 p.

49. Jiao S, Ren H, Li Y, Zhou J, Duan C, et al. (2013) Differential regulation of IGF-I and IGF-II gene expression in skeletal muscle cells. Mol Cell Biochem 373: 107–113. doi:10.1007/s11010-012-1479-4.

50. Bassom F (2009) Mini encyclopedia of chicken breeds & care: a color directory of the most popular breeds and their care. Bufflao, N.Y.; Richmond Hill, Ont.: Firefly Books.

51. Eyras E, Reymond A, Castelo R, Bye JM, Camara F, et al. (2005) Gene finding in the chicken genome. BMC Bioinformatics 6: 131. doi:10.1186/1471-2105-6-131.

52. Elferink MG, Megens H-J, Vereijken A, Hu X, Crooijmans RPMA, et al. (2012)

Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken

Breeds. PLoS ONE 7: e32720. doi:10.1371/journal.pone.0032720.

4th Chapter

Whole-genome scan for detection of signatures of recent selection

Mahmood Gholami[1], Christian Reimer[1], Malena Erbe[1], Rudolf Preisinger[2], Annett Weigend[3], Steffen Weigend[3], Tim-Matthias Strom[4] and Henner Simianer[1]

1 Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Göttingen, Albrecht-Thaer-Weg 3, 37075 Göttingen
2 LOHMANN Tierzucht GMBH, Am Seedeich 9-11, P.O.Box 460, 27454 Cuxhaven
3 Institute of Farm Animal Genetics (ING), Friedrich-Loeffler-Institut (FLI), Höltystraße 10, 31535 Neustadt
4 Institut für Humangenetik, Helmholtz Zentrum München, Ingolstädter Landstr. 1, D-85764 Neuherberg

Paper in Progress

Abstract

We used SNPs extracted from pool sequence data to detect selection signatures. Over 30 million raw SNPs in 43 pools from 43 different breeds (including three wild species: *G. g. gallus*, *G. g. spadiceus* and *G. varius*) were used in this study. After filtering, more than 22 million SNPs remained for analysis. The breeds were studied for selection signatures in three contrasts (i.e. skin color, egg shell color, and toe number). F_{ST} (Wright's fixation index) was calculated between the two groups in each contrast, whereas heterozygosity was obtained for each group. The averages of F_{ST} and heterozygosity values were calculated for overlapping windows within each contrast. Comparisons of averages of F_{ST} and heterozygosity between overlapping windows were employed to detect selection signatures, this was done to improve the power and reliability of detection. A total of eight regions (in all contrast) were detected as selective signatures using the combination of both methods. Annotation of selection signature regions revealed one gene (*BCO2*) and three QTL corresponding to skin color and egg shell color, respectively. In this study we demonstrate that the use of sequence data with a larger number of populations and combination of different methods of selection signature detection can improve the power of detection.

Introduction

Natural or artificial selection on a beneficial mutation causes a local reduction of genetic variation up- and downstream of the beneficial mutation, leaving special patterns of DNA behind (Smith and Haigh, 1974). Linked variation is thus swept through the population along with the beneficial mutation; a process referred to as a ‘‘selective sweep’’ (Akey, 2009). The study of such patterns can provide valuable insights into genomic regions harboring interesting genes. Hence it may also help to understand the mechanisms that lead to the differentiation of various genotypes and the corresponding influenced phenotypes during selection.

Genomic tests for selection signature detection can be distinguished by their corresponding summary statistics. A number of these statistics are used to detect signatures of hitchhiking events (Qanbari and Simianer, 2014). For example Pooled Heterozygosity (H_P) was used to detect reduction of local variability (Rubin et al., 2010), maximum of composite likelihood ratio (CLR) was employed to detect allele frequencies spectrum deviation (Nielsen et al.,

2005), and relative extended haplotype homozygosity (rEHH) was used to detect specific linkage disequilibrium pattern (Sabeti et al., 2002). An alternative method for selection signature detection from genomic data is based on genetic diversity between populations (Lewontin and Krakauer, 1973). Innan and Kim (2008) and Yi *et al.* (2010) showed that among recently diverged populations, methods based on genetic diversity have more statistical power for the detection of selection signatures. Wright's fixation index, F_{ST}, is a useful index for genetic differentiation between populations (Wright, 1949). Regions under diversifying selection should exhibit larger divergence among population than neutral loci (high F_{ST}), while regions under uniform balancing selection in all population should be less differentiated (low F_{ST}). Compared to the hitchhiking approach, the F_{ST} method focuses on a different selection scenario: diversifying local selection instead of population-wide positive selection (Hermisson, 2009). Ascertainment bias can have a great impact on genomic tests for selection, particularly on the tests based on frequency spectrum and population differentiation, such as F_{ST} (Clark et al., 2005). The best solution would be a combination of the full genome sequence data instead of low resolution genotyping assays (Qanbari et al., 2014), and a larger number of populations (Gholami et al., n.d.).

The growing genomic resources, the relatively short reproduction time and the existence of several inbred lines, together with strong agricultural interest makes chicken an excellent model for studying the signatures of selection under artificial breeding conditions (Brown et al., 2003). Chicken meat and eggs are one of the major protein sources for humans; in addition, chicken has been a popular model organism for at least 100 years (Stern, 2005). A better understanding of the chicken genome can enhance the use of chicken as a model organism for biomedical research (Burt, 2007) and can lead to the identification of genes or regions of the genome that are associated with beneficial traits. Several groups have studied selection signatures in chicken by focusing on methods based on reduced local variability. Rubin *et al.* (2010) scanned the genome by calculating Pooled Heterozygosity (H_P) of 9 different lines for selection signature using pool sequence data. Johansson *et al.* (2010) used 60k SNP chip to scan the genome of two chicken lines which had 9-fold difference in body weight due to artificial selection and used observed heterozygosity as a method for selection signature detection. Elferink *et al.* (2012) calculated H_P in order to scan the genome of 67 lines for selection signature using 58K SNP chip data. To our knowledge our studies (Gholami et al., 2014, nd) were so far the only ones applying methods based on genetic diversity between populations for selection signature detection in chickens, in that the genome

of 3 commercial egg-layers were scanned for recent selection using 1 million SNPs with F_{ST} based methods.

In this study, we used pool sequences (10 individuals per pool) from 43 different chicken breeds. The 43 chicken breeds consisted of 3 commercial breeds (White leghorn, White Rock and Rhode Island Red) and 40 non-commercial breeds (including three wild species: *G. g. gallus*, *G. g. spadiceus* and *G. varius*). We calculated F_{ST} and H_P for several comparison groups (i.e. skin color, egg shell color, and toe number) and used a combination of both methods to verify putative selection signatures. Our dataset is unprecedented with regards to the combination of number of genotyped individuals and the applied marker densities, which can provide important information about the genomic regions which have been under selection with specific traits in chicken.

Material and methods

Animals, data collection and filtering

This study was conducted on a set of 43 breeds, comprising of 3 commercial breeds (White leghorn, White Rock and Rhode Island Red) and 40 non-commercial breeds (including three wild species: *G. g. gallus*, *G. g. spadiceus* and *G. varius*). A more detailed list of breeds is presented in Table 1.

DNA was isolated from 10 female individuals in each breed using a phenol/chloroform method for DNA isolation (Sambrook and Russell, 2001). DNA was isolated from whole blood collected from the wing vein using EDTA as anticoagulant. White leghorn individuals were sequenced individually with 8x coverage and subsequently a virtual pool was created. For other breeds a DNA pool was made from 10 individuals and sequenced with 20x coverage on the NGS-Platform.

Raw sequence data was aligned to the Galgal4 reference genome (International Chicken Genome Sequencing Consortium, 2004) using BWA (Li and Durbin, 2009). Then, the alignments were sorted and duplications were marked with Picard tools (“Picard,” n.d.). Variants (SNPs, indels and SNV) were called using GATK (DePristo et al., 2011; McKenna et al., 2010). Only SNPs on autosomal chromosomes were used for this study, therefore we

removed indels, SNV and all the variation on none-autosomal chromosomes (sex chromosomes, mitochondria chromosome and linkage groups). We also filtered for unreliable SNPs by removing SNPs in clusters with > 5 SNPs in 20 basepairs, with BaseQualityRankSum < -5.18 or > 3.78, MappingQualityRankSum < -9.75 or > 4.35, ReadPosRankSum < -2.9, FisherStrand values > 12, Mapping Quality < 30, and Depth of Coverage < 622X or > 1117X. To pass subsequent genotype filtering an individual required a genotyping quality > 20 and a pool needed a coverage > 4 reads at this position. A total of 22,942,249 SNPs remained after all the filtering.

Table 1. Name, abbreviation, number of individuals, skin color, egg color and number of toes for each breed used in this study.

Breed English name	Abbreviation	Skin color	Egg color	# of Toes
Antwerpener Bartzwerge wachtelfarbig	AB	white	white to cream	4
Aseel red mottled	AS	yellow	brown	4
Bergische Kraeher (Bergische Crower)	BK	white	brown	4
Booted Bantam millefleur	FG	white	white to cream	4
Brahma gold	BH	yellow	brown	4
Castella (black)	KA	white	brown	4
Cochin black	CS	yellow	brown	4
Deutsche Lachshuehner lachsfarbig (German Faverolles salmon)	DL	white	brown	5
East Friesian Gulls silver pencilled	OM	white	brown	4
Gallus gallus gallus	GG	white	unknown/unsure	4
Gallus gallus spadiceus	GS	white	unknown/unsure	4
Gallus varius	GV	white	unknown/unsure	4
German Bantam gold partridge	DG	white	white to cream	4
Hollaender Weisshaube blau-gesaeumt	HO	white	white	4
Italiener rebhuhnhalsig (Leghorn brown)	IT	yellow	white	4
Japanese Bantam	CG	yellow	white to cream	4
Japanese Bantam black mottled	CW	yellow	white to cream	4
Ko Shamo black-red	KG	yellow	brown	4
Krueper schwarz (Creeper black)	KS	white	white	4
Krueper weiss (Creeper white)	KW	white	white	4
Leghorn Linie R11	LE	yellow	white	4
Malay	MA	yellow	yellow	4
Marans copper black	MR	white	dark brown	4
New Hampshire L68	NH	yellow	brown	4
Oh Shamo black	SH	yellow	brown	4
Ohiki silver duckwing	OH	yellow	brown	4
Orloff red spangled	OF	yellow	white to cream	4
Orpington buff	OR	white	brown	4
Pekin Bantam white	ZC	yellow	brown	4
Poland any colour	PA	white	white	4
Rhode Island Red	RI	yellow	brown	4
Rosecomb Bantam black	BA	white	white to cream	4
Rumpless Araucana black	AR	yellow	turquoise	4
Sebright Bantam silver	SB	white	white to cream	4
Silkies white	SE	black	brown	5
Sumatra black	SA	white	white to yellow	4
Sundheimer light	SN	white	brown	4
Toutenkou black breasted red	TO	yellow	white to yellow	4
Westphalian Chicken silver	WT	white	white	4
White Leghorn	WL	yellow	white	4
White Rock	WR	yellow	brown	4
Wyandotte white	WY	yellow	brown	4
Yokohama red saddled white	YO	yellow	yellow	4

Genome wide scans

To detect selection signatures, we defined several contrasts: skin color, egg color and number of toes. In order to build the contrasts, animals with similar and trustable phenotype were pooled together into two groups in each contrast, and only loci with missing genotyping less than 50% were included. Information on each contrast is available in Table 2.

Table 2. Information on each contrast.

Contrast	Group1 (# of breeds)	Group2 (# of breeds)
Skin color	White (21)	Yellow (21)
Egg color	White (16)	Brown (18)
Toe number	4 toes (41)	5 toes (2)

In each contrast, Wright's fixation index (F_{ST}) (Wright, 1949) was calculated between two groups. The z-transformed expected heterozygosity (zH_E) (Rubin et al., 2010) was calculated for both groups in each contrast. The z-transformation produced comparability of several breed pools with differing average heterozygosity within the same framework, because quintile-based thresholds can be applied more easily on normalized values (Rubin et al., 2010). Both measures (F_{ST} and zH_E) were subsequently summarized in 40 kb windows with an overlap of 50%. Based on the genome-wide distribution of F_{ST} (zH_E), a threshold cutting off the upper (lower) 1% was used for defining extreme values.

Annotation

A region was qualified as a putative selective sweep if it had a size bigger than 100 kb and was detected (exceeded the threshold) by F_{ST} and zH_E in one of the groups of contrast (e.g. White or Yellow group in Skin color contrast). All putative sweeps within 500 kb of each other were grouped together. For all putative sweep groups gene annotations, QTL annotations and pathway annotations were completed. Gene and QTL annotations were done with BEDTools (Quinlan and Hall, 2010) using Ensembl database (Flicek et al., 2013) and animal QTL database (Hu et al., 2013), respectively. Gene enrichment analysis was done with Fisher's exact test (Fisher, 1922) for all annotated genes in each contrast separately. Pathways and gene ontologies with $p \leq 0.05$ were identified as being under selection. KEGG database (Kanehisa et al., 2012) was used for pathways and gene ontologies analysis.

Results and Discussion

Overview

Average F_{ST} value in the egg color contrast (0.029, SE=0.00001) was higher than in the skin color (0.018, SE= 0.000004) and toe number (0.008, SE=0.000005) contrasts. This finding supports the fact that egg color is a quantitative trait which is controlled by a number of genes (regions) (Hutt, 1949). Consequently, although the breeds in one group may show similarity in their phenotype, in fact their phenotype can be controlled by different regions. Detection of selective sweeps for polygenic traits (e.g. egg color) is much more difficult than for monogenic traits (e.g. skin color). This has also been observed in other studies, for example in detection of coat coloring pattern in cattle (Qanbari et al., 2014), and dark brown feather plumage color in chicken (Gholami et al., n.d.). In the following sections we report and discuss the result of each contrast in more detail.

Selective sweep detection

Defining a window size of 40 kb with an overlap of 50% was a decision motivated by previous studies (Qanbari et al., 2012; Rubin et al., 2010). We also summarized our measures (F_{ST} and zH_E) in 100 kb windows with an overlap of 50% which gave nearly identical results (result not shown); therefore we presented the 40 kb results.

Although the outlier approach is an effective method for identifying the region under selection lacking known phenotypes (Qanbari et al., 2011), an outlier signal is not necessarily synonymous with regions being under selection (Akey, 2009). Solo analysis of our methods (F_{ST} and zH_E) resulted in large amount of polymorphisms as signatures of selection, which does not appear realistic (Nuzhdin and Turner, 2013). Thus, we used a combination of both methods to evaluate a region as a putative selective sweep. As demonstrated in Table 3, there is a significant reduction in detection rate when the two methods are combined, compare to using each method separately.

Table 3. Regions detected as under selection with F_{ST}, H_E and combination of two methods in different contrast.

Contrast	F_{ST}	H_E	Combined
Egg color	228	96	3
Skin Color	231	101	2
Toe number	201	95	3

Skin color

Based on the F_{ST} and zH_E values distribution and size of the signal (bp), we detected two regions as putative selective sweep in skin color contrast (as shown in Figure 1). A more detailed list of these regions is presented in Table 4. We detected a clear signal on chromosome 24 which harbors *BCO2*, i.e. beta-carotene oxygenase 2 (Eriksson et al., 2008). Inhibition of *BCO2* causes yellow skin phenotype in chicken. This region was selected extensively in chicken since there is a strong consumer preference for the yellow skin phenotype in many geographic markets (Castañeda et al., 2005). This region was detected in several studies as a region under selection (Elferink et al., 2012; Rubin et al., 2010). Therefore we used it as a proof of principle for the method of selection signature detection in this study. Heterozygosity declines in this region in the yellow skin group (Figure 2), which indicates fixation of this region in yellow skin chicken. Our results support the findings of Rubin *et al.* (2010) and Elferink *et al.* (2012).

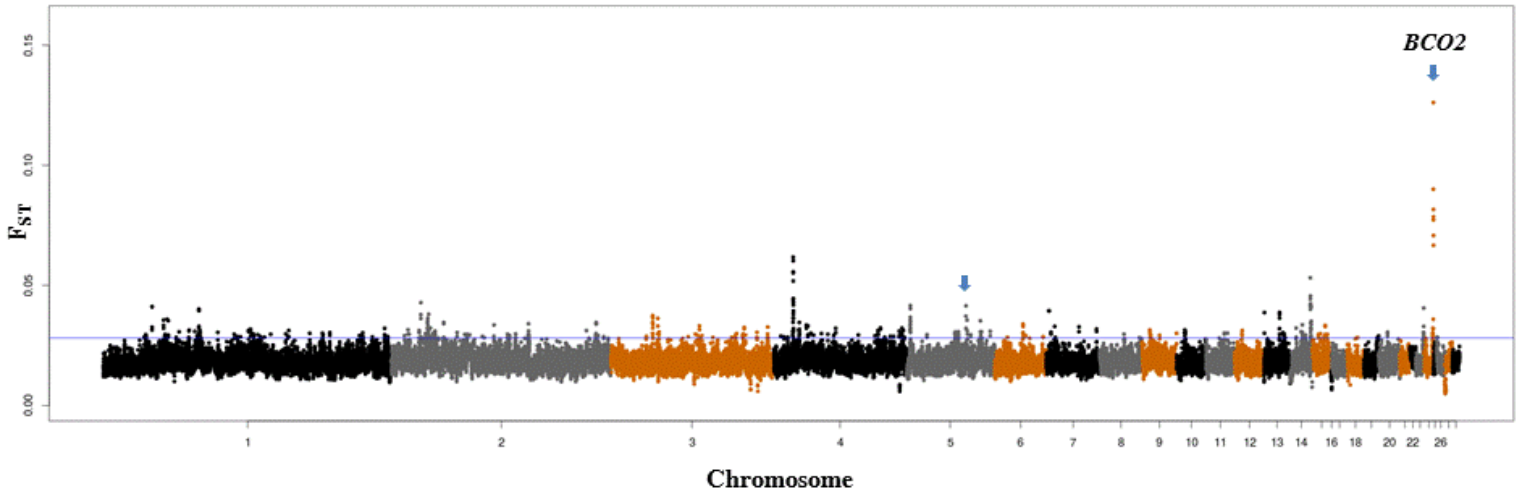

Figure 1. F_{ST}-values of 40 Kbp windows (with 50% overlap) in skin color contrast. Blue line indicates the upper 1% of FST distribution. Blue arrows indicate the regions that qualified as putative selective sweep after evaluation.

Table 4. Regions detected as putative sweeps in skin color contrast.

Chr	Start (bp)	End (bp)	Average F_{ST} value (SE)	Average White zH_E (SE)	Average Yellow zH_E (SE)
5	2,220,000	2,440,000	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
24	6,060,000	6,300,000	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)

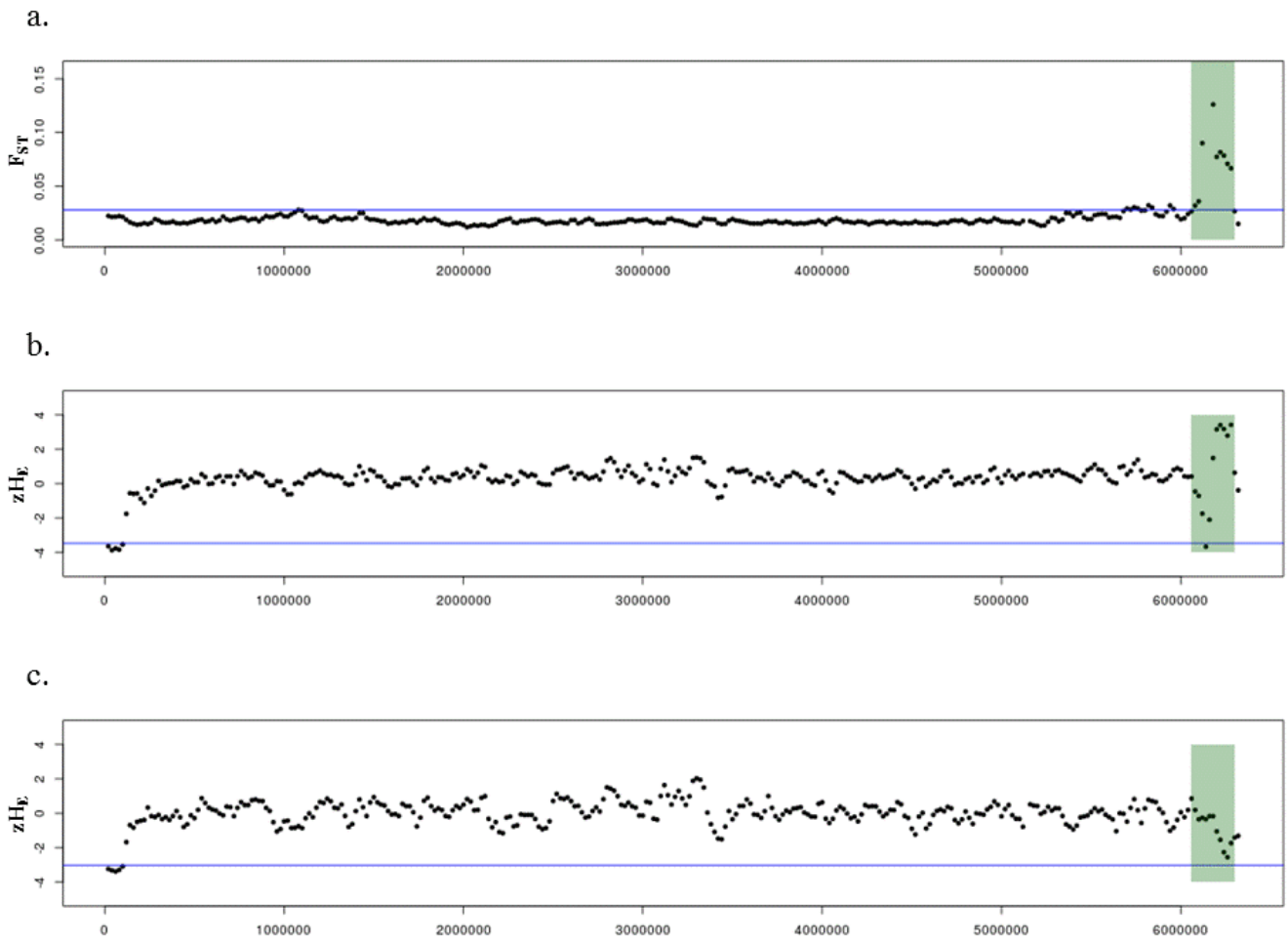

Figure 2. Plot of chromosome 24 for skin color contrast. Blue line indicates the 1% upper (lower) distribution of F_{ST} (zH_E), for, (a) F_{ST}, (b) heterozygosity in yellow skin group and (c) heterozygosity in white skin group. Green box indicates the location of the putative sweep around *BOC2*.

We did not detect any gene with direct association to skin color in the region on chromosome 5. However, we annotated several genes and pathways that may be related to skin color indirectly (in both regions). A complete list of all genes and pathways with more details is available in Table S1 and Table 5, respectively.

Table 5. Pathways annotated in putative sweeps in skin color contrast.

Description	# Genes annotated	Genes of pathways (%)*	P-Value
structural constituent of eye lens	2	15.4	0.001
positive regulation of cytokinesis	2	12.5	0.001
sensory perception of smell	2	6.9	0.003
synaptic vesicle	2	4.8	0.005
identical protein binding	4	1.2	0.009
locomotory behavior	2	2.9	0.014
dendrite	2	2.6	0.017
ribosome	2	2.4	0.020
axon	2	2.1	0.026
external side of plasma membrane	2	1.6	0.040

*Percentage of the genes of the pathway which were among the annotated genes.

Egg color

Using the above mentioned sweep evaluation method, we detected three regions as putative sweeps in the egg color contrast (Figure 3). A more detailed list of these regions is presented in Table 6. On chromosome 11, we detected a selection signature that harbors three QTL which are associated with egg shell color; egg shell lightness (Sasaki et al., 2004), egg shell yellowness and egg shell redness (Sasaki et al., 2004). Heterozygosity value around this region indicates selection on brown layers for this region (Figure 4). In addition, several QTL associated with production in chicken were annotated in the same region (chr11: 18,120,000-18,780,000) such as: egg number and body weight. A list of all annotated QTL in egg color contrast is available in Table 7. We also annotated *MC1R* (melanocortin 1 receptor) on chromosome 11, which is associated with plumage color diversity (Guo et al., 2010). A list of all genes and pathways detected in egg color contrast is available in Table S2 and S3, respectively.

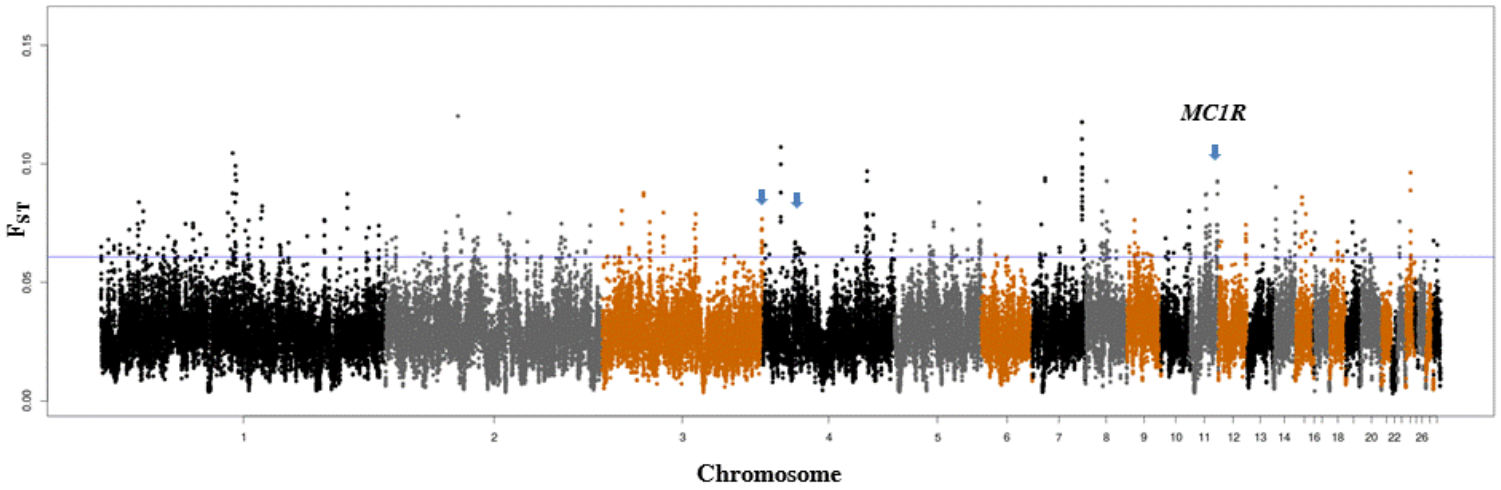

Figure 3. F_{ST}-values of 40 Kbp windows (with 50% overlap) in egg color contrast. Blue line indicates the upper 1% of F_{ST} distribution. Blue arrows indicate the regions that qualified as putative selective sweep after evaluation.

Table 6. Regions detected as putative sweeps in egg color contrast.

Chr	Start (bp)	End (bp)	Average F_{ST} value (SE)	Average White zH_E (SE)	Average Brown zH_E (SE)
3	109,480,000	109,740,000	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
4	1,860,000	21,980,000	0.06 (0.0003)	-0.008 (0.061)	-0.111 (0.052)
11	18,120,000	18,780,000	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)

Table 7. QTL annotated in putative sweeps in egg color contrast.

Chr	Start (bp)	End (bp)	QTL	Average F_{ST} value (SE)	Average White zH_E (SE)	Average Brown zH_E (SE)
11	18970494	18970594	Abdominal fat percentage	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Abdominal fat weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Age at first egg	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Antibody titer to SRBC antigen	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Body weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Breast color	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Breast muscle percentage	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Breast muscle weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Carcass weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Egg number	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Egg shell lightness	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Egg shell redness	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Egg shell strength	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Egg shell thickness	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Egg shell weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Egg shell yellowness	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Feather pecking	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Feed conversion ratio	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Femur bending strength	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Growth	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18455194	18498934	Heart weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Spleen weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	19280000	Thigh bone weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17620000	17756128	Thigh meat-to-bone ratio	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Thigh muscle percent	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Thigh muscle weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Thymus weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18970494	18970594	Yolk weight	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)

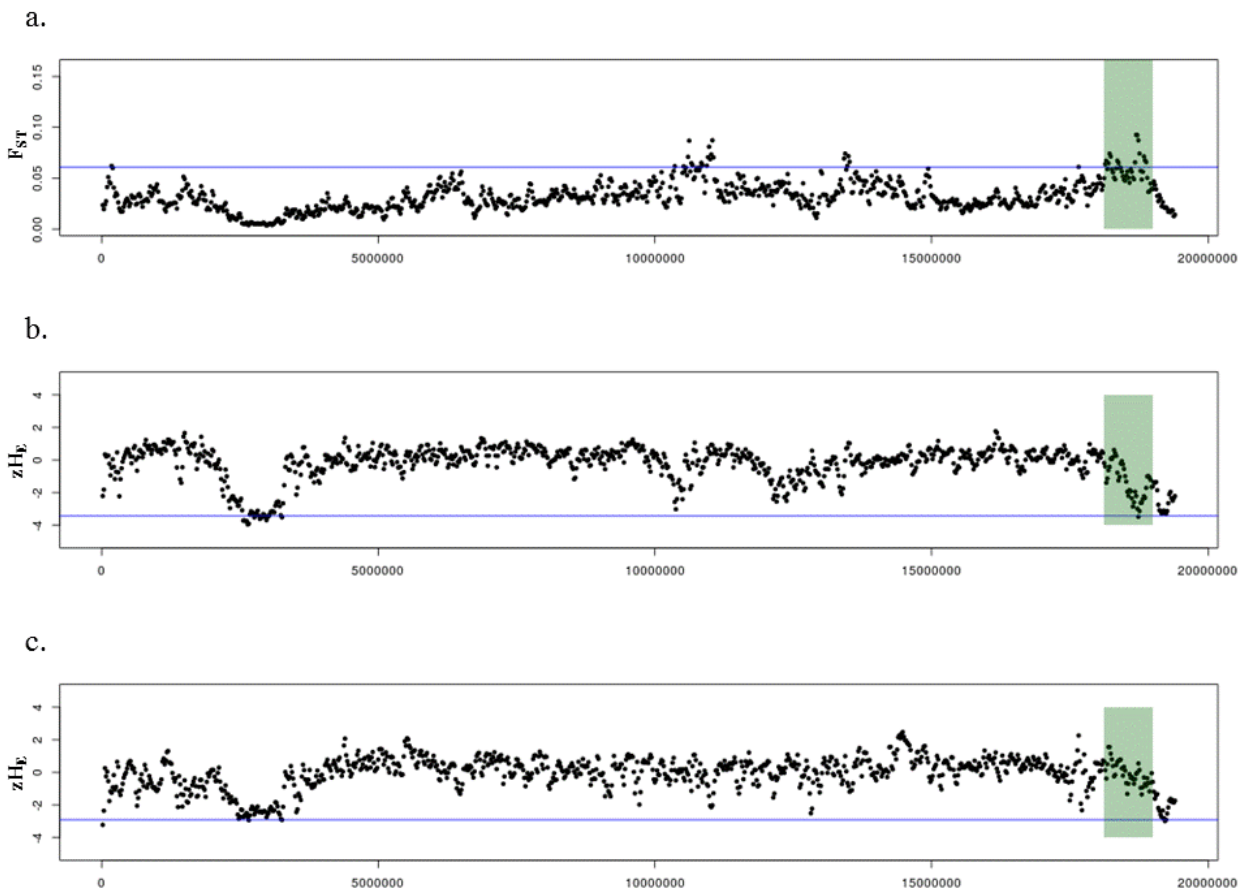

Figure 4. Plot of chromosome 11 for egg color contrast. Blue line indicates the upper (lower) of F_{ST} (zH_E) distribution, for, (a) F_{ST}, (b) heterozygosity in brown egg group and (c) heterozygosity in white egg group. Green box indicates the location of the putative sweep around QTL associated to egg color.

Number of toes

The toe number contrast had the lowest average F_{ST} (0.008) compared to the other two contrasts. This can be partially due to the fact that in the toe number contrast only two breeds had five toes (41 breeds with 4 toes) which results in a lack of diversity in the 5-toe group. Nevertheless we detected several signals with high F_{ST} in this contrast (Figure 5); however after evaluation (sweep size and detection by heterozygosity) only three regions were selected as putative sweeps. A list of these sweeps with more details is available in Table 8.

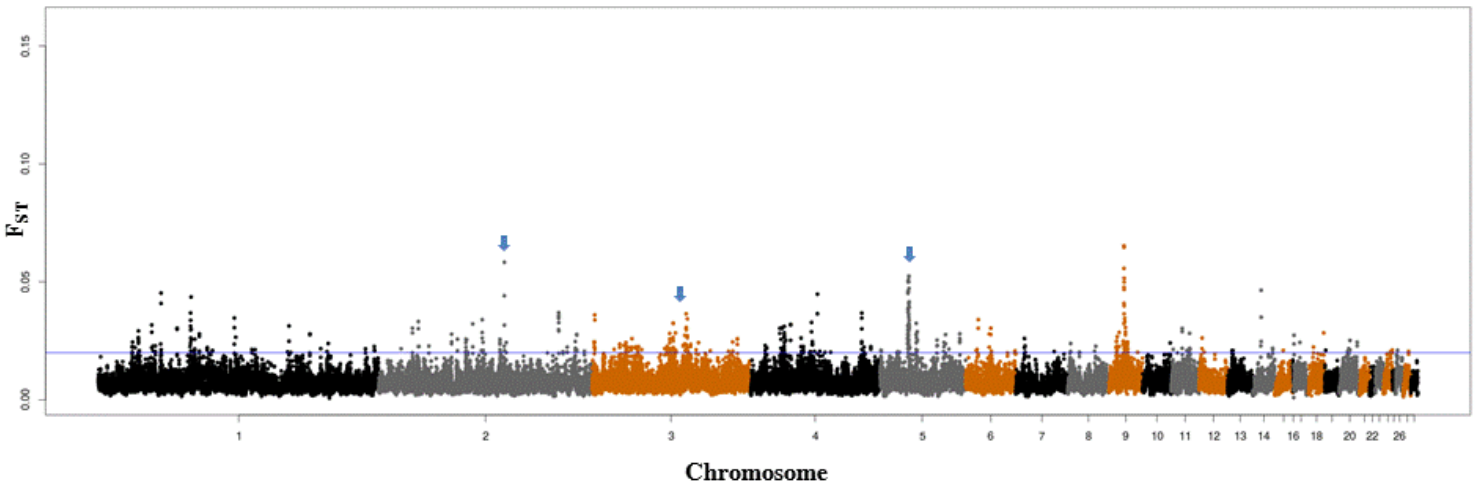

Figure 5. F_{ST}-values of 40 Kbp windows (with 50% overlap) in toe number contrast. Blue line indicates the upper 1% of F_{ST} distribution. Blue arrows indicate the regions that qualified as putative selective sweep after evaluation.

Table 8. Regions detected as putative sweeps in toe number contrast.

Chr	Start (bp)	End (bp)	Average F_{ST} value (SE)	Average 5 Toe zH_E (SE)	Average 4 Toe zH_E (SE)
2	125,440,000	125,600,000	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
3	1,420,000	1,540,000	0.02 (0.0001)	-0.61 (0.112)	-0.93 (0.120)
5	19,200,000	20,320,000	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)

In contrast to our expectation we did not detect any genes associated with toe number in any of these regions, which could partly be due to insufficient knowledge about these regions (Eyras et al., 2005). However, the region detected on chromosome 2 could be the same region mapped for polydactyly by Pitel et al., (2000). We detected several pathways as pathways under selection which may be indirectly associated with the number of toes in chicken. Lists of all genes and pathways annotated in the toe number contrast is available in Table S4 and S5, respectively.

Outlook

Further evaluation of these results using other methods of selection signature detection, e.g. sweep finder (Nielsen et al., 2005), may give a better resolution of the discussed selection signature detection method. On the other hand, an analysis of the data with methods that account for hierarchical structure, e.g. FLK (Bonhomme et al., 2010), may give new insight on selection signature in chicken breeds.

Conclusion

In conclusion we were able to identify several putative selective signature regions with genes or QTL corresponding to phenotype diversity in our contrasts. As a proof of principle, we found the *BCO2* gene as the top hit in the skin color contrast. Additionally, we detected three QTL associated with egg color in the respective contrast. Previous studies on selection signatures for egg color failed to detect regions corresponding to egg color (Gholami et al., 2014, n.d.). Ultimately, further analysis and optimization of the discussed evaluation approach might give a higher reliability to our method of sweep detection.

Supporting Information

Supplementary Table 1. List of genes for selective sweeps detected in skin color contrast.

Chr	Start (bp)	End (bp)	Gene	Description	Average F_{ST} value (SE)	Average White zH_E (SE)	Average Yellow zH_E (SE)
5	1720000	1846316	E2F8	E2F transcription factor 8	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
5	1856963	2042983	NAV2	neuron navigator 2	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
5	2045882	2049582	DBX1	developing brain homeobox 1 (DBX1), mRNA.	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
5	2129859	2187426	PRMT3	protein arginine methyltransferase 3 (PRMT3), mRNA.	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
5	2204842	2231619	SLC6A5	solute carrier family 6 (neurotransmitter transporter), member 5	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
5	2243249	2526700	NELL1	NEL-like 1 (chicken)	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
5	2760044	2792484	ANO5	Gallus gallus anoctamin 5 (ANO5), mRNA.	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
5	2813098	2843784	SLC17A6	solute carrier family 17, member 6	0.03 (0.001)	-0.59 (0.331)	-2.99 (0.128)
24	5574123	5587498	DDX6	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6 (DDX6), mRNA.	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5590150	5595172	CXCR5	chemokine (C-X-C motif) receptor 5 (CXCR5), mRNA.	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5601464	5612457	BCL9L	B-cell CLL/lymphoma 9-like	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5657150	5658883	RPS25	ribosomal protein S25	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5658963	5660793	TRAPPC4	trafficking protein particle complex 4 (TRAPPC4), mRNA.	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5661175	5668655	SLC37A4	solute carrier family 37 (glucose-6-phosphate transporter), member 4	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5669909	5683409	HYOU1	hypoxia up-regulated protein 1 precursor	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5684173	5705399	PHLDB1	pleckstrin homology-like domain, family B, member 1	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5732694	5757116	DRD2	dopamine receptor D2 (DRD2), mRNA.	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5769723	5780497	ANKK1	ankyrin repeat and kinase domain containing 1	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	5786864	5800976	TTC12	tetratricopeptide repeat domain 12	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)

24	5803420	5884198	NCAM1	neural cell adhesion molecule 1 isoform 1 precursor	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6128919	6131573	PTS	6-pyruvoyltetrahydropterin synthase	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6133777	6154197	BCO2	beta-carotene oxygenase 2	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6163149	6165998	IL18	interleukin 18 (interferon-gamma-inducing factor) (IL18), mRNA.	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6167374	6171382	SDHD	succinate dehydrogenase complex, subunit D, integral membrane protein (SDHD), nuclear gene encoding mitochondrial protein, mRNA.	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6179094	6188311	DLAT	dihydrolipoamide S-acetyltransferase	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6190722	6217912	DIXDC1	DIX domain containing 1	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6230555	6234006	HSPB2	heat shock 27kDa protein 2 (HSPB2), mRNA.	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6235233	6238779	CRYAB	Alpha-crystallin B chain	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)
24	6243258	6247754	FDXACB1	ferredoxin-fold anticodon binding domain containing 1	0.09 (0.019)	-0.97 (0.314)	-2.05 (0.612)

Supplementary Table 2. Pathways annotated in putative sweeps in Egg color contrast.

Description	# Genes anotated	Genes of pathways (%)*	P-Value
protein domain specific binding	4	4.7	0.002
response to oxidative stress	3	5.5	0.005
myelination	2	9.5	0.007
cell-cell adherens junction	2	8.7	0.009
heme binding	3	4.1	0.011
electron carrier activity	2	6.7	0.015
response to stimulus	2	6.5	0.016
cytokinesis	2	6.3	0.017
skeletal system morphogenesis	2	6.1	0.018

axon	3	3.1	0.022
cell periphery	2	5.4	0.022
cytoplasm	25	1.0	0.023
G-protein coupled receptor signaling pathway	6	1.7	0.025
SH3 domain binding	2	4.9	0.027
Rab GTPase binding	2	4.5	0.031
protein localization	2	4.4	0.032
vesicle-mediated transport	3	2.7	0.034
G-protein coupled receptor activity	5	1.7	0.037
single organismal cell-cell adhesion	2	3.8	0.042
neuropeptide signaling pathway	2	3.6	0.048

Supplementary Table 3. List of genes for selective sweeps detected in Egg color contrast.

Chr	Start (bp)	End (bp)	Gene	Description	Average F_{ST} value (SE)	Average White zH_E (SE)	Average Brown zH_E (SE)
3	108988859	109073478	RUNX2	runt-related transcription factor 2	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109252462	109301586	CLIC5	chloride intracellular channel 5	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109323128	109327586	ENPP4	ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative) (ENPP4), mRNA.	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109338624	109369202	RCAN2	regulator of calcineurin 2	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109423489	109444248	CYP39A1	cytochrome P450, family 39, subfamily A, polypeptide 1	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)

3	109447789	109456799	TDRD6	tudor domain containing 6	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109455352	109469298	PLA2G7	platelet-activating factor acetylhydrolase precursor	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109476672	109477638	TAS2R7	taste receptor, type 2, member 7 (TAS2R7), mRNA.	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109487074	109499681	MEP1A	meprin A, alpha (PABA peptide hydrolase) (MEP1A), mRNA.	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109500726	109534703	GPR116	G protein-coupled receptor 116	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109558089	109588428	TNFRSF21	tumor necrosis factor receptor superfamily, member 21 (TNFRSF21), mRNA.	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109620064	109688541	CD2AP	CD2-associated protein	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109727730	109741746	OPN5	opsin 5 (OPN5), mRNA.	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	109787390	109871532	C3H6ORF138	chromosome 3 open reading frame, human C6orf138 (C3H6ORF138), mRNA.	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
3	110190487	110240000	EVA1A	eva-1 homolog A (C. elegans)	0.07 (0.001)	-1.04 (0.083)	0.38 (0.256)
4	21360000	21376386	RXFP1	relaxin/insulin-like family peptide receptor 1	0.06 (0.0003)	-0.008 (0.061)	-0.111 (0.052)
4	21379597	21396998	ETFDH	electron-transferring-flavoprotein dehydrogenase (ETFDH), nuclear gene encoding mitochondrial protein, mRNA.	0.06 (0.0003)	-0.008 (0.061)	-0.111 (0.052)
4	21395222	21405790	PPID	peptidylprolyl isomerase D	0.06 (0.0003)	-0.008 (0.061)	-0.111 (0.052)
4	21425882	21457823	FNIP2	folliculin interacting protein 2	0.06 (0.0003)	-0.008 (0.061)	-0.111 (0.052)
4	21509901	21683609	RAPGEF2	Rap guanine nucleotide exchange factor (GEF) 2	0.06 (0.0003)	-0.008 (0.061)	-0.111 (0.052)
4	22250826	22480000	FSTL5	follistatin-like 5	0.06 (0.0003)	-0.008 (0.061)	-0.111 (0.052)
11	17844834	17846138	CIDEC	cell death-inducing DFFA-like effector c (CIDEC), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17890559	17892594	IL17C	interleukin 17C	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17893778	17895490	CYBA	cytochrome b-245, alpha polypeptide	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)

11	17895873	17898830	MVD	mevalonate (diphospho) decarboxylase	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17900343	17906187	RNF166	ring finger protein 166 (RNF166), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17907682	17911287	CTU2	cytosolic thiouridylase subunit 2 homolog (S. pombe)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17911235	17930659	PIEZO1	piezo-type mechanosensitive ion channel component 1	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17932050	17935844	CDT1	chromatin licensing and DNA replication factor 1	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17936905	17939727	APRT	adenine phosphoribosyltransferase (APRT), transcript variant 1, mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17944300	17988567	GALNS	galactosamine (N-acetyl)-6-sulfate sulfatase	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17988612	17991639	TRAPPC2L	trafficking protein particle complex 2-like	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	17999325	18017927	CBFA2T3	core-binding factor, runt domain, alpha subunit 2; translocated to, 3 (CBFA2T3), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18039023	18076346	ACSF3	acyl-CoA synthetase family member 3	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18078995	18142059	ANKRD11	ankyrin repeat domain 11	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18174075	18177448	AFG3L2	AFG3-like AAA ATPase 2	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18179332	18183587	RPL13	ribosomal protein L13 (RPL13), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18185943	18190546	CPNE7	copine VII	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18195279	18198549	DPEP1	dipeptidase 1 (renal)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18202288	18205788	CHMP1A	chromatin modifying protein 1A (CHMP1A), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18207482	18211559	CDK10	cyclin-dependent kinase 10 (CDK10), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18212171	18213607	SPATA2L	spermatogenesis associated 2-like	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18213719	18218083	VPS9D1	VPS9 domain containing 1	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)

11	18225501	18257747	FANCA	Fanconi anemia, complementation group A	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18260412	18267088	SPIRE2	spire-type actin nucleation factor 2	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18267772	18286192	TCF25	transcription factor 25 (basic helix-loop-helix)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18287876	18288821	MC1R	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor) (MC1R), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18290820	18292152	TUBB3	tubulin, beta 3 class III (TUBB3), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18293265	18296749	DEF8	differentially expressed in FDCP 8 homolog (mouse)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18293265	18296749	DEF8	differentially expressed in FDCP 8 homolog (mouse)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18297074	18298237	DBNDD1	dysbindin (dystrobrevin binding protein 1) domain containing 1	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18300554	18308190	GAS8	growth arrest-specific 8 (GAS8), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18310976	18316940	CDH3	cadherin 3, type 1, P-cadherin (placental)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18317571	18326276	CDH1	cadherin 1, type 1, E-cadherin (epithelial) (CDH1), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18317571	18326276	CDH1	cadherin 1, type 1, E-cadherin (epithelial) (CDH1), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18355335	18357463	HAS3	hyaluronan synthase 3	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18358801	18360418	CHTF8	CTF8, chromosome transmission fidelity factor 8 homolog (S. cerevisiae) (CHTF8), transcript variant 2, mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18360465	18365214	CIRH1A	cirrhosis, autosomal recessive 1A (cirhin)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18365412	18370296	SNTB2	syntrophin, beta 2 (dystrophin-associated protein A1, 59kDa, basic component 2)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18378689	18380054	NIP7	nuclear import 7 homolog (S. cerevisiae) (NIP7), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18383994	18391771	TERF2	telomeric repeat binding factor 2 (TERF2), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18396031	18407915	CYB5B	outer mitochondrial membrane cytochrome b5 (CYB5B), nuclear gene encoding mitochondrial protein, mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)

11	18415929	18474957	NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive (NFAT5), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18482779	18484568	NQO1	NAD(P)H dehydrogenase	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18487654	18519315	WWP2	WW domain containing E3 ubiquitin protein ligase 2 (WWP2), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	18520345	18526403	PSMD7	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	19150948	19161411	DHX38	DEAH (Asp-Glu-Ala-His) box polypeptide 38 (DHX38), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	19162774	19166255	DHODH	dihydroorotate dehydrogenase (quinone) (DHODH), nuclear gene encoding mitochondrial protein, mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	19166597	19173788	IST1	increased sodium tolerance 1 homolog (yeast)	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	19192363	19194436	ATXN1L	ataxin 1-like	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	19195951	19237680	AP1G1	adaptor-related protein complex 1, gamma 1 subunit (AP1G1), mRNA.	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)
11	19238987	19268628	PHLPP2	PH domain and leucine rich repeat protein phosphatase 2	0.06 (0.002)	0.168 (0.192)	-1.154 (0.33)

Supplementary Table 4. List of genes for selective sweeps detected in toe number contrast.

Chr	Start (bp)	End (bp)	Gene	Description	Average FST value (SE)	Average 5 Toe zHE (SE)	Average 4 Toe zHE
2	125273981	125333683	TRIQK	triple QxxK/R motif containing	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125604623	125621196	FAM92A1	family with sequence similarity 92, member A1	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125623602	125630097	RBM12B	RNA binding motif protein 12B (RBM12B), mRNA.	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125683594	125690722	PDP1	pyruvate dehyrogenase phosphatase catalytic subunit 1	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125754666	125783019	CDH17	cadherin 17, LI cadherin (liver-intestine) (CDH17)	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125802914	125811572	GEM	GTP binding protein overexpressed in skeletal muscle (GEM), mRNA.	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)

2	125844097	125903394	RAD54B	RAD54 homolog B (S. cerevisiae)	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125877034	125887013	FSBP	fibrinogen silencer binding protein	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125907864	125917414	RNF151	ring finger protein 151	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125925065	125952171	KIAA1429	KIAA1429	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125956619	125988739	ESRP1	epithelial splicing regulatory protein 1	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	125956619	125988739	ESRP1	epithelial splicing regulatory protein 1	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	126052839	126076741	DPY19L4	dpy-19-like 4 (C. elegans)	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
2	126082913	126100000	NULLWERT	cHz-cadherin (LOC414835), mRNA.	0.03 (0.0006)	1.97 (0.059)	-0.21 (0.142)
3	920000	929281	FANCL	Fanconi anemia, complementation group L (FANCL), mRNA.	0.02 (0.0001)	-0.61 (0.112)	-0.93 (0.120)
3	1831201	1886847	BCL11A	B-cell CLL/lymphoma 11A (zinc finger protein) (BCL11A), mRNA.	0.02 (0.0001)	-0.61 (0.112)	-0.93 (0.120)
3	1970600	1986711	PAPOLG	poly(A) polymerase gamma	0.02 (0.0001)	-0.61 (0.112)	-0.93 (0.120)
3	2009614	2040000	REL	v-rel reticuloendotheliosis viral oncogene homolog (avian) (REL), mRNA.	0.02 (0.0001)	-0.61 (0.112)	-0.93 (0.120)
5	18700000	18704538	COMMD9	COMM domain containing 9	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)
5	18727686	18766719	PRR5L	proline rich 5 like	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)
5	18776382	18793694	TRAF6	TNF receptor-associated factor 6, E3 ubiquitin protein ligase	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)
5	18812000	18817045	RAG2	V(D)J recombination-activating protein 2	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)
5	19803070	19804987	LRRC4C	leucine rich repeat containing 4C	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)
5	20702879	20716727	API5	apoptosis inhibitor 5 (API5), mRNA.	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)
5	20718590	20772087	TTC17	tetratricopeptide repeat domain 17	0.03 (0.001)	0.64 (0.119)	-1.25 (0.142)

Supplementary Table 5. Pathways annotated in putative sweeps in toe number contrast.

Description	# Genes anotated	Genes of pathways (%)*	P-Value
protein ubiquitination	2	1.4	0.036
positive regulation of I-kappaB kinase	2	2.2	0.015
positive regulation of interleukin-12 biosynthetic	2	40.0	0.000
cytoplasmic side of plasma membrane	2	8.3	0.001
B cell differentiation	2	8.3	0.001
T cell differentiation	2	8.7	0.001
homophilic cell adhesion	2	3.3	0.007
ubiquitin protein ligase binding	3	2.8	0.001
ubiquitin-protein ligase activity	3	1.8	0.005
nucleic acid binding	4	0.9	0.015
metal ion binding	7	0.7	0.003

References

Akey, J.M., 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19, 711–722. doi:10.1101/gr.086652.108

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., SanCristobal, M., 2010. Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. Genetics 186, 241–262. doi:10.1534/genetics.110.117275

Brown, W.R.A., Hubbard, S.J., Tickle, C., Wilson, S.A., 2003. The chicken as a model for large-scale analysis of vertebrate gene function. Nat. Rev. Genet. 4, 87–98. doi:10.1038/nrg998

Burt, D.W., 2007. Emergence of the Chicken as a Model Organism: Implications for Agriculture and Biology. Poult. Sci. 86, 1460–1471. doi:10.1093/ps/86.7.1460

Castañeda, M.P., Hirschler, E.M., Sams, A.R., 2005. Skin pigmentation evaluation in broilers fed natural and synthetic pigments. Poult. Sci. 84, 143–147.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., Nielsen, R., 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15, 1496–1502. doi:10.1101/gr.4107905

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498. doi:10.1038/ng.806

Elferink, M.G., Megens, H.-J., Vereijken, A., Hu, X., Crooijmans, R.P.M.A., Groenen, M.A.M., 2012. Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds. PLoS ONE 7, e32720. doi:10.1371/journal.pone.0032720

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., Jensen, P., Andersson, L., 2008. Identification of the Yellow Skin Gene Reveals a Hybrid Origin of the Domestic Chicken. PLoS Genet 4, e1000010. doi:10.1371/journal.pgen.1000010

Eyras, E., Reymond, A., Castelo, R., Bye, J.M., Camara, F., Flicek, P., Huckle, E.J., Parra, G., Shteynberg, D.D., Wyss, C., Rogers, J., Antonarakis, S.E., Birney, E., Guigo, R., Brent, M.R., 2005. Gene finding in the chicken genome. BMC Bioinformatics 6, 131. doi:10.1186/1471-2105-6-131

Fisher, R.A., 1922. On the Interpretation of $\chi 2$ from Contingency Tables, and the Calculation of P. J. R. Stat. Soc. 85, 87–94. doi:10.2307/2340521

Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Garcia-Giron, C., Gordon, L., Hourlier, T., Hunt, S., Juettemann, T., Kahari, A.K., Keenan, S., Komorowska, M., Kulesha, E., Longden, I., Maurel, T., McLaren, W.M., Muffato, M., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H.S., Ritchie, G.R.S., Ruffier, M., Schuster, M., Sheppard, D., Sobral, D., Taylor, K., Thormann, A.,

Trevanion, S., White, S., Wilder, S.P., Aken, B.L., Birney, E., Cunningham, F., Dunham, I., Harrow, J., Herrero, J., Hubbard, T.J.P., Johnson, N., Kinsella, R., Parker, A., Spudich, G., Yates, A., Zadissa, A., Searle, S.M.J., 2013. Ensembl 2013. Nucleic Acids Res. 41, D48–D55. doi:10.1093/nar/gks1236

Gholami, M., Erbe, M., Gärke, C., Preisinger, R., Weigend, A., Weigend, S., Simianer, H., 2014. Population Genomic Analyses Based on 1 Million SNPs in Commercial Egg Layers. PLoS ONE 9, e94509. doi:10.1371/journal.pone.0094509

Gholami, M., Reimer, C., Erbe, M., Preisinger, R., Weigend, A., Weigend, S., Servin, B., Simianer, H., n.d. Genome scan for selection in structured layer chicken populations exploiting linkage disequilibrium information.

Guo, X.L., Li, X.L., Li, Y., Gu, Z.L., Zheng, C.S., Wei, Z.H., Wang, J.S., Zhou, R.Y., Li, L.H., Zheng, H.Q., 2010. Genetic variation of chicken MC1R gene in different plumage colour populations. Br. Poult. Sci. 51, 734–739. doi:10.1080/00071668.2010.518408

Hermisson, J., 2009. Who believes in whole-genome scans for selection? Heredity 103, 283–284. doi:10.1038/hdy.2009.101

Hu, Z.-L., Park, C.A., Wu, X.-L., Reecy, J.M., 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 41, D871–D879. doi:10.1093/nar/gks1150

Hutt, F.B., 1949. II. Variation in eggs, in: Genetics of the Fowl. McGRAW-HILL book company, INC., New York.

Innan, H., Kim, Y., 2008. Detecting Local Adaptation Using the Joint Sampling of Polymorphism Data in the Parental and Derived Populations. Genetics 179, 1713–1720. doi:10.1534/genetics.108.086835

International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716. doi:10.1038/nature03154

Johansson, A.M., Pettersson, M.E., Siegel, P.B., Carlborg, Ö., 2010. Genome-Wide Effects of Long-Term Divergent Selection. PLoS Genet 6, e1001188. doi:10.1371/journal.pgen.1001188

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, D109–114. doi:10.1093/nar/gkr988

Kim, Y., Stephan, W., 2002. Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. Genetics 160, 765–777.

Lewontin, R.C., Krakauer, J., 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74, 175–195.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 25, 1754–1760. doi:10.1093/bioinformatics/btp324

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. doi:10.1101/gr.107524.110

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005. Genomic scans for selective sweeps using SNP data. Genome Res. 15, 1566–1575. doi:10.1101/gr.4252305

Nuzhdin, S.V., Turner, T.L., 2013. Promises and limitations of hitchhiking mapping. Curr. Opin. Genet. Dev. 23, 694–699. doi:10.1016/j.gde.2013.10.002

Picard [WWW Document], n.d. URL http://picard.sourceforge.net/ (accessed 8.8.14).

Pitel, F., Bergé, R., Coquerelle, G., Crooijmans, R.P., Groenen, M.A., Vignal, A., Tixier-Boichard, M., 2000. Mapping the naked neck (NA) and polydactyly (PO) mutants of the chicken with microsatellite molecular markers. Genet. Sel. Evol. GSE 32, 73–86. doi:10.1051/gse:2000107

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G., Simianer, H., 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics 12, 318. doi:10.1186/1471-2164-12-318

Qanbari, S., Strom, T.M., Haberer, G., Weigend, S., Gheyas, A.A., Turner, F., Burt, D.W., Preisinger, R., Gianola, D., Simianer, H., 2012. A High Resolution Genome-Wide Scan for Significant Selective Sweeps: An Application to Pooled Sequence Data in Laying Chickens. PLoS ONE 7, e49525. doi:10.1371/journal.pone.0049525

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T.M., Fries, R., Nielsen, R., Simianer, H., 2014. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. PLoS Genet 10, e1004148. doi:10.1371/journal.pgen.1004148

Qanbari, S., Simianer, H., 2014. Mapping signatures of positive selection in the genome of livestock. Livest. Sci. 166, 133–143. doi:10.1016/j.livsci.2014.05.003

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. doi:10.1093/bioinformatics/btq033

Rubin, C.-J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., Andersson, L., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591. doi:10.1038/nature08832

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837. doi:10.1038/nature01140

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. CSHL Press.

Sasaki, O., Odawara, S., Takahashi, H., Nirasawa, K., Oyamada, Y., Yamamoto, R., Ishii, K., Nagamine, Y., Takeda, H., Kobayashi, E., Furukawa, T., 2004. Genetic mapping of quantitative trait loci affecting body weight, egg character and egg production in F2 intercross chickens. Anim. Genet. 35, 188–194. doi:10.1111/j.1365-2052.2004.01133.x

Smith, J.M., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23, 23–35. doi:10.1017/S0016672300014634

Stern, C.D., 2005. The chick; a great model system becomes even greater. Dev. Cell 8, 9–17. doi:10.1016/j.devcel.2004.11.018

Wright, S., 1949. The Genetical Structure of Populations. Ann. Eugen. 15, 323–354. doi:10.1111/j.1469-1809.1949.tb02451.x

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., Zheng, H., Liu, T., He, W., Li, K., Luo, R., Nie, X., Wu, H., Zhao, M., Cao, H., Zou, J., Shan, Y., Li, S., Yang, Q., Asan, Ni, P., Tian, G., Xu, J., Liu, X., Jiang, T., Wu, R., Zhou, G., Tang, M., Qin, J., Wang, T., Feng, S., Li, G., Huasang, Luosang, J., Wang, W., Chen, F., Wang, Y., Zheng, X., Li, Z., Bianba, Z., Yang, G., Wang, X., Tang, S., Gao, G., Chen, Y., Luo, Z., Gusang, L., Cao, Z., Zhang, Q., Ouyang, W., Ren, X., Liang, H., Zheng, H., Huang, Y., Li, J., Bolund, L., Kristiansen, K., Li, Y., Zhang, Y., Zhang, X., Li, R., Li, S., Yang, H., Nielsen, R., Wang, J., Wang, J., 2010. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science 329, 75–78. doi:10.1126/science.1190371

5th Chapter

General discussion

Preface

This thesis studies the effects of genome resolution and different methods in selection signature detection. These two factors were studied in **chapters 2**, **3** and **4**. In this chapter first an overview of the results from **chapters 2**, **3** and **4** is provided, afterwards the challenges in selection signature detection and possible solution are discussed.

Overview

Chicken, notably the commercial lines, has a small effective population size (Qanbari et al., 2010) and therefore genetic drift is high (Nielsen and Slatkin, 2013). Detection of selection signatures in (effectively) small populations with high drift is difficult due to an overlap of regions that differentiate by drift and selection. Additionally, it is shown that existence of extreme bottlenecks during domestication, which is the case in chicken (Rao, 2007), causes the absence of background variation, which precludes the chance of selection signature detection (Hamblin et al., 2006). Small effective population size and absence of background variation might be partly the reason that regions known to be associated with egg color were not detected in **chapters 2** and **3**. A solution to this problem is to use a larger number of populations to minimize the risk of random creation of a systematic pattern of differentiation between breed groups. Detection of regions harboring three QTL associated to egg color in **chapter 4** may be the outcome of including a higher number of populations. Nonetheless regions with strong evidence of selection were detected in **chapters 2** and **3**, such as regions associated to meat production in dual purpose breeds in both chapters as well as a region underlying the dark brown mutational phenotype in chickens' plumage (Gunnarsson et al., 2011) in **chapters 3**.

Overall detection of selection signature of monogenic traits (e.g. skin color, feather plumage color) is much easier than for polygenic traits (e.g. egg shell color, egg number) as discussed in **chapters 2** and **3,** where detection of regions associated to egg shell color fails, and partly in **chapter 4**, where the power of detection is much lower compare to monogenic traits (i.e. skin color). Other studies also have detected monogenic traits with higher power than other traits; skin color gene detection in chicken (Elferink et al., 2012; Rubin et al., 2010) and coat coloring pattern in cattle (Qanbari et al., 2014) are two such examples. One of the major concerns in selection signature detection in chicken is that most production traits in chicken (e.g. number of eggs, body weight and feed conversion) are polygenic traits (Crawford, 1990).

However, construction of a clear contrast (e.g. several layers vs. several non-commercials) may aid the detection. As a result of the clear contrast several regions associated to production traits were detected in **chapter 2** which harbored genes such as *NCOA1*, *IGFII* and *POMC*, which are associated to egg production, growth and feed conversion, respectively. Moreover, a clear contrast combined with a large population size lead to detection of a putative region that determines brownish egg shell color, which is a polygenic trait, as discussed in **chapter 4** (Hutt, 1949).

Scan resolution

In this thesis two sets of data were used. The first set has one million SNPs obtained from three SNP-chips during the development of the high density Affymetrix 600K SNP arrays for chicken (Kranis et al., 2013) which is used in **chapter 2** and **3**. The second set has 23 million SNPs obtained from pool sequence data which is used in **chapter 4**. Several studies have remarked the limited resolution and ascertainment bias when using SNP-chip data, which is caused by the process used to discover SNPs (Nielsen, 2000; Wakeley et al., 2001) and the criteria applied in the selection of SNPs to be put on commercial SNP chips. However, due to the comparably small size of chicken genome (1.2 Gbp) and the high amount of SNPs (~ one million) in our dataset, there was no lack of resolution in **chapter 2** and **3**. Figure 1 demonstrates the SNPs distribution all over the genome in both data sets. On average the density of SNPs was 20 times higher in pool sequence data in comparison to the one-million SNP-chip.

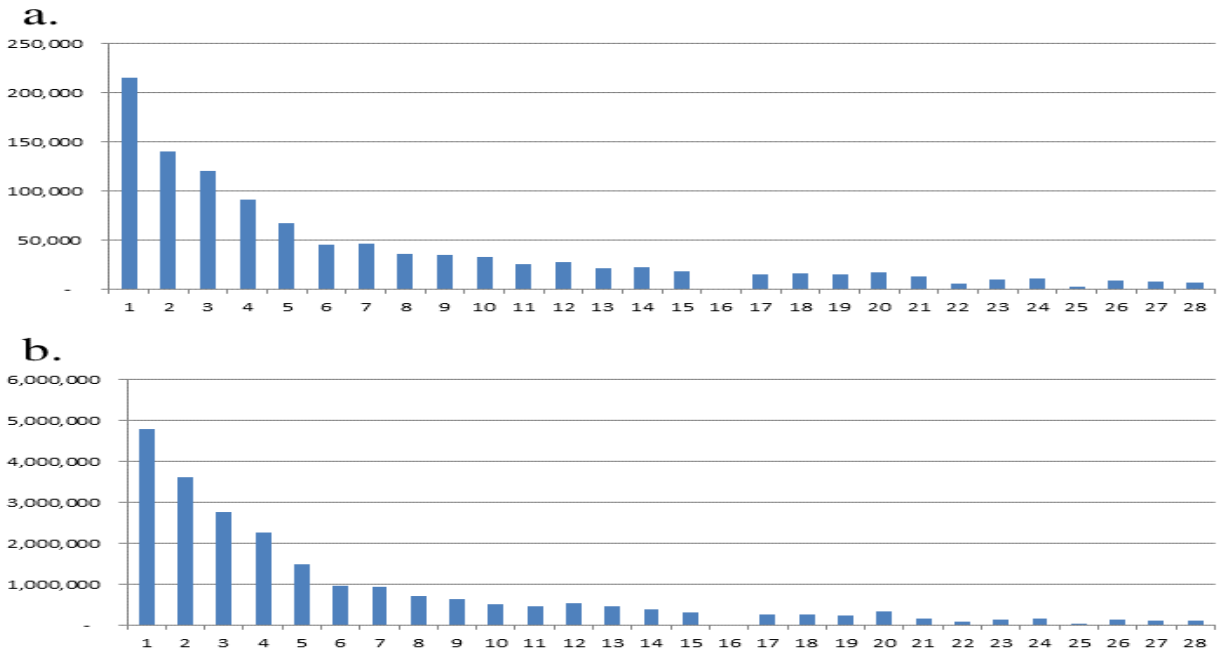

Figure 1. SNPs distribution in (a) 1 million SNP dataset and (b) pool sequence data.

Tests based on frequency spectrum and population differentiation such as F_{ST} are more subject to ascertainment bias (Clark et al., 2005) than other tests such as LD based tests (Tang et al., 2007). Since our analysis with two different datasets was on different sets of populations, comparison of the effect of resolution between these two dataset is not accurate. Therefore, to evaluate the effect of high resolution scan, F_{ST} analysis of skin color (from **chapter 4**) was run on a subset of SNPs on chromosome 24 (from **chapter 2** and **3**). From 166,443 SNPs that existed in the pool dataset on chromosome 24, 8,873 common SNPs that were in the one-million SNPs dataset were selected. As demonstrated in Figure 2, a slight reduction of noise is visible with pool data. However, both sets of SNPs detected the region with *BCO2* gene with high power. In closure, it can be claimed that extreme high resolution scan has a lower effect on power of detection compare to higher number of population and clear contrast. It is important to note that our first data set (one-million) is considered to represent high resolution in chicken as well.

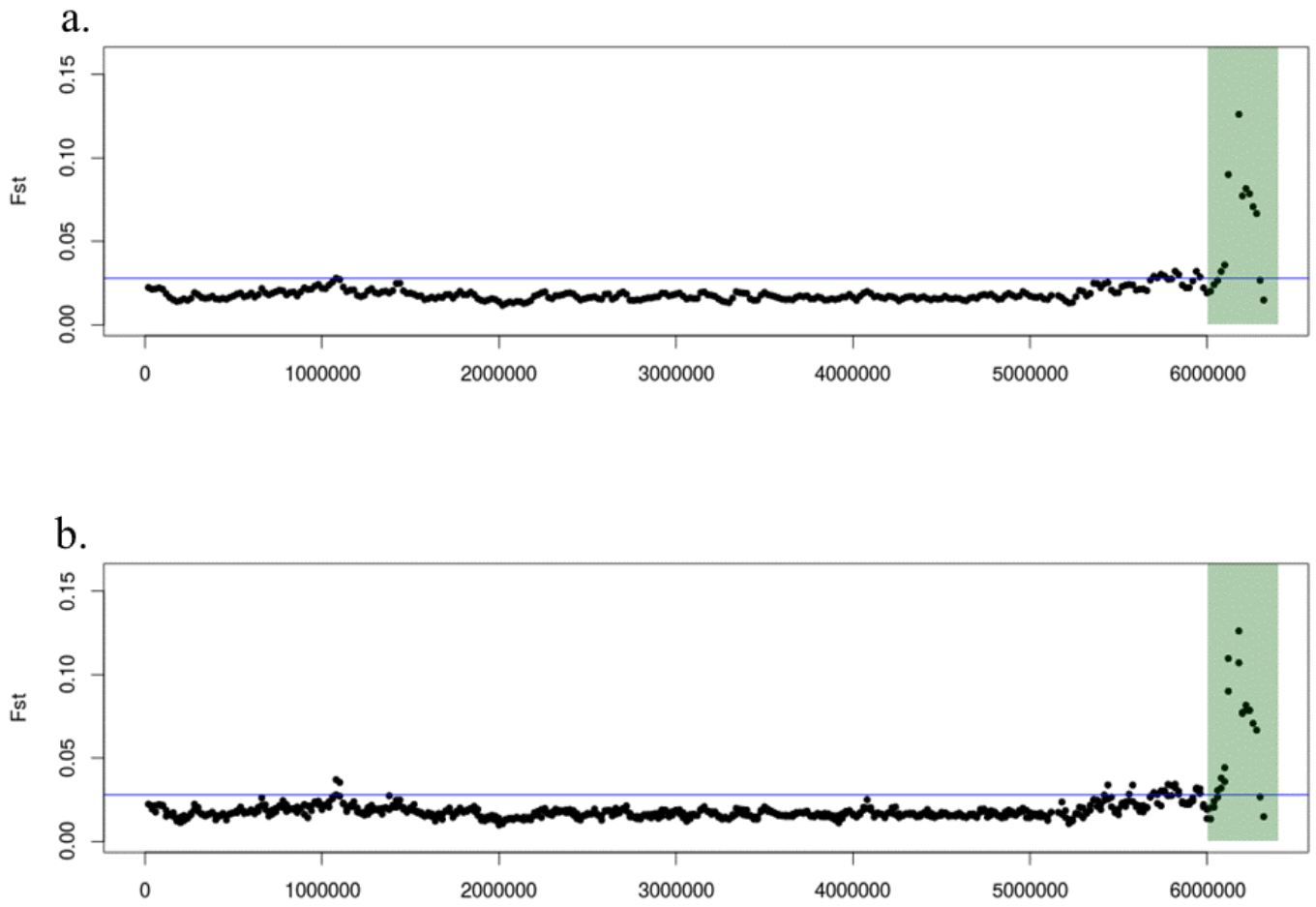

Figure 2. F_{ST} plot of Chromosome 24 for Skin color. Blue line indicates the upper 1 % of F_{ST} distribution, for (a) pool sequence data (b) one-million SNPs data. Green box indicates the location of the putative sweep around *BCO2*.

Method of detection

Four methods were used in this thesis for selection signature detection: F_{ST}, FLK, hapFLK and H_E, three of which (F_{ST}, FLK and hapFLK) were based on inter-population statistics. Several studies have shown that between recently diverged populations methods based on inter-population statistics have more power for detection of selection signatures than intra-population statistics (Innan and Kim, 2008; Yi et al., 2010). On the contrary, an advantage of intra-population methods (e.g. H_E) is the capability of pinpointing the population that has been the target of selection. However in **chapter 3**, re-estimation of the branch lengths in the population tree with local allele or haplotype cluster frequencies was used to identify the branch lengths that seem significantly larger than the branches of the whole genome tree in order to pinpoint selected populations (for details read methods in **chapter 3**).

F_{ST} is the most widely used inter-populations statistic which has been extensively used for detection of selection signatures. Several studies have used F_{ST} statistic for detection of natural selection in humans (Amato et al., 2009; Barendse et al., 2009; Porto-Neto et al., 2013) or impact of artificial selection on domesticated animals (Johansson et al., 2010; Vaysse et al., 2011; Yang et al., 2014). It is demonstrated in **chapter 4** that F_{ST} has considerable power in detection of selection on monogenic traits by detection of *BCO2* gene which causes the yellow skin color in chicken (Eriksson et al., 2008). Additionally, by detection of regions associated with egg color in **chapter 4**, it is demonstrated that F_{ST} is capable to detect selection on polygenic traits when a clear contrast with several populations is formed. On the other hand, lack of clear contrast in **chapter 2** resulted in lower power of detection. The average F_{ST} values of contrasts presented in **chapter 2** are higher than contrasts in **chapter 4** (Table 1), i.e. a higher degree of differentiation is observed between contrasts in **chapter 2** compared to **chapter 4**. Therefore, possibly a higher risk of a systematic pattern of differentiation exists in contrasts discussed in **chapter 2** in comparison with contrasts in **chapter 4**. However, several regions associated to meat production and one region associated to egg production in the contrast of commercial layers vs. non-commercial breeds were detected in **chapter 2**. Consequently, it can be claimed that F_{ST} is a powerful statistic for detection of selection signatures between groups of population. Nonetheless F_{ST} statistic has its disadvantages.

Table 1. Average F_{ST} with standard error (SE) in contrast of chapter 2 and 4.

	Contrast	F_{ST} (SE)
Chapter 2	Layers vs. Non-commercial	0.116 (0.00005)
	White layers vs. Brown layers	0.257 (0.00009)
Chapter 4	Yellow skin vs. White skin	0.018 (0.000004)
	White layers vs Brown layers	0.029 (0.00001)
	4 toe vs. 5 toe	0.008 (0.000005)

A major concern with F_{ST} is that it implicitly assumes that populations have the same effective size (*Ne*) and were derived independently from an ancestral population in a star-like phylogeny. When this is not true F_{ST} might produce false positive signals, similar to the well-known effects of cryptic structure in genome-wide association studies (Price et al., 2010). Therefore in **chapter 3** FLK (Bonhomme et al., 2010) a statistic that deals with *Ne* variation and historical branching of populations was used. Using simulation data Bonhomme *et al.* (2010) showed that FLK is indeed more powerful than F_{ST} for a given false positive rate. A much lower number of selection signatures were detected with FLK in **chapter 3** (7.9%) compared to the F_{ST} in **chapter 2** (Figure 3), this might be partly due to higher false positive error with the F_{ST} method. Although that genes or QTL associated to egg production or egg color were not detected with the FLK method in **chapter 3**, several QTL and genes associated to body weight and carcass trait were detected. Lack of diversity may be the main obstacle in detection of regions associated to egg production or egg color in **chapter 3**. Using the FLK method on the pool sequence data from **chapter 4** might give a better insight into the regions under selection for egg color, mainly due to the high number of population in the pool sequence dataset.

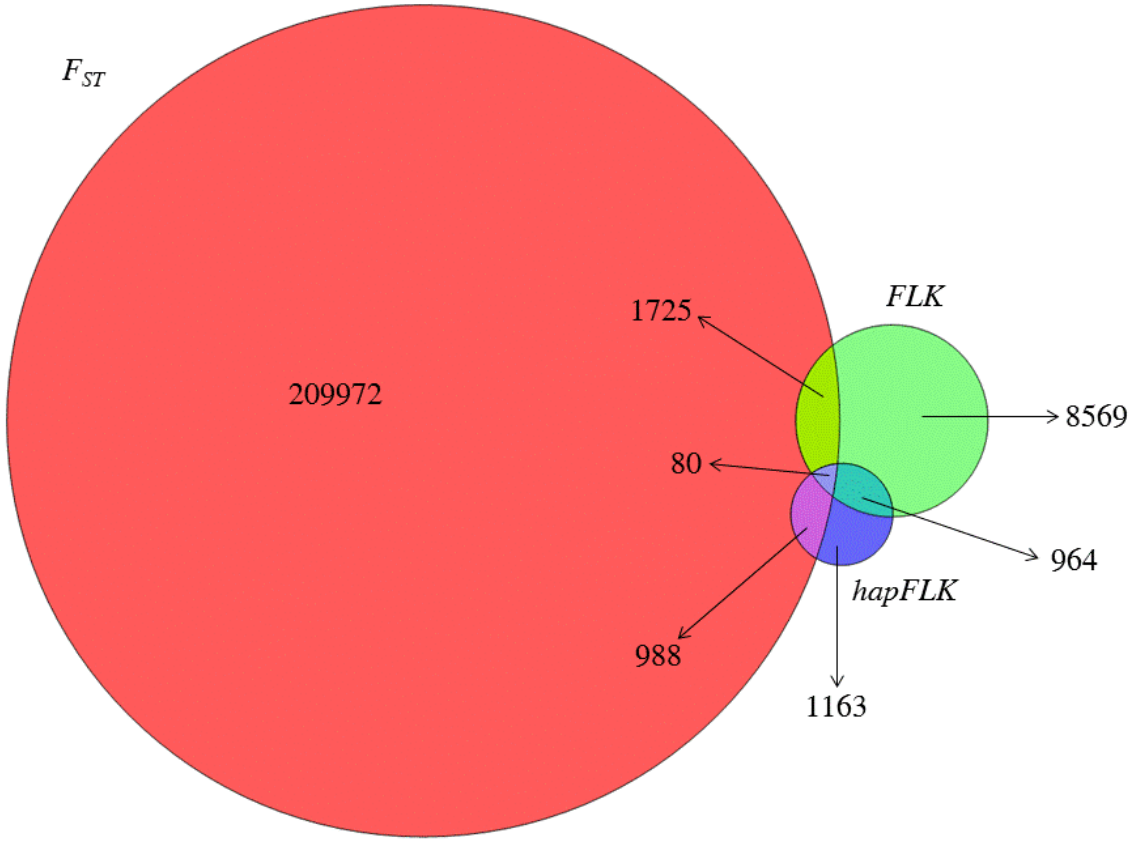

Figure 3. Venn diagram of overlapping SNPs identified as under selection, with F_{ST} in chapter 2, FLK and hapFLK in chapter 3.

Haplotype diversity contains useful information for selection signature detection (Sabeti et al., 2007), and additionally, SNP ascertainment bias has less impact on haplotype based differentiation analyses compared to single site differentiation analyses (Browning and Weir, 2010). Most of the haplotype differentiation scans do not account for the possibility of hierarchical structure between populations (Qanbari and Simianer, 2014). Hence, in **chapter 3** hapFLK (Fariello et al., 2013), a haplotype based extension of the FLK statistic that accounts for both hierarchical structure and haplotype information was used. Simulation studies showed that the hapFLK statistic has more power in detection of soft sweeps, incomplete sweeps and sweeps occurring in several populations in comparison to other F_{ST}-like methods (Fariello et al., 2013). In this analysis with hapFLK several genes and QTL associated with growth and carcass traits were detected. Although no gene or QTL associated with egg shell color or egg production was detected, a region that harbors a gene associated to dark brown phenotype in chickens' plumage was detected. Further analysis showed that this region was under selection in the only breed that has dark brown feather in the first dataset (for details see discussion in **chapter 3**).

Several studies have used heterozygosity for selection signature detection in domesticated animals (Quilez et al., 2011; Rubin et al., 2012), along with few studies which used pooled heterozygosity for selection signature detection in chicken (Elferink et al., 2012; Qanbari et al., 2012; Rubin et al., 2010). In **chapter 4**, heterozygosity was used to evaluate putative regions that have been detected as region under selection with the F_{ST} method. In addition, use of heterozygosity gave the possibility to pinpoint the population that was under selection.

Outlier method

The outlier approach is an effective and widely used method for identification of genes under selection lacking known phenotypes (Narum and Hess, 2011). This method identifies loci with the most unusual pattern of variation as targets of selection (typically the tails of the distribution -top/bottom 1%- of a test statistic); however an outlier signal is not necessarily synonymous with regions being under selection (Akey, 2009). Several studies have argued the necessities of proper statistical tests revealing statistical significance for candidate selection signals (Kelley et al., 2006; Teshima et al., 2006). Yet due to the difficulties of deriving a null distribution of the test statistic in these tests (Qanbari and Simianer, 2014), researchers often focus on the outlier approach and avoid specifying a statistical model (Akey et al., 2002). In this thesis, we attempted to develop a statistical approach to determine signature of selection. In **chapter 2** a permutation test with 100 replications was conducted in order to determine a null threshold. Results showed that the F_{ST} distribution under randomization is much lower than the observed distribution of F_{ST}, which corresponded to a threshold 10 times lower than the threshold that was used (top/bottom 1%). In **chapter 3** a gamma distribution was fitted to the hapFLK observed distribution, using the minimum distance estimation method (Zhou et al., 2011), which is robust to outliers. This method was used to estimate p-values of selection signatures which suggested a false discovery rate (FDR) of 10-20% in this analysis. Using the gamma distribution in the evaluation of selection signatures resulted in detection of only a few regions as putative selective sweeps. These regions did not harbor any gene or QTL associated to any production traits in chicken (result not discussed in this thesis).

In Figure 4 regions that have been detected as region under selection with F_{ST}, FLK and hapFLK methods are demonstrated. As it is demonstrated in Figure 4, using the outlier method in F_{ST} analysis of **chapter 2** resulted in detection of selection signatures in all chromosomes of chicken. Detection of thousands of polymorphisms as putative candidate under selection, which is the case in many studies, does not appear realistic (Nuzhdin and

Turner, 2013). Therefore, in **chapter 3** a much stricter threshold was used in FLK and hapFLK analysis (top/bottom 0.05%).

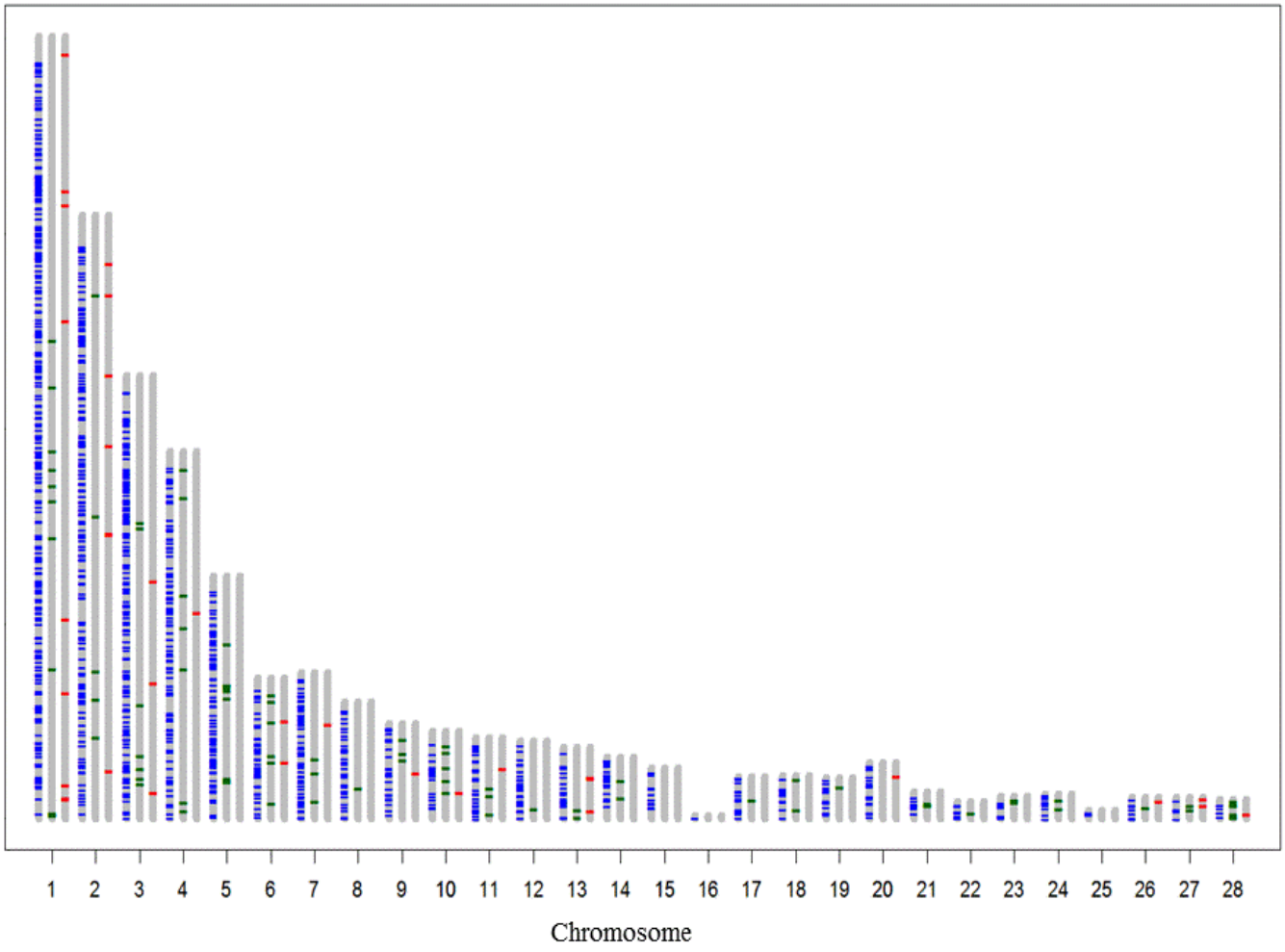

Figure 4. Regions detected as putative selection signatures. Blue, green and red lines indicate regions detected by F_{ST} with 1% threshold, FLK and hapFLK with 0.05% threshold, respectively.

Despite all the drawbacks of the outlier approach, using it has resulted in many interesting candidate targets of selection (Qanbari and Simianer, 2014). In this thesis these were regions associated to meat production in dual purpose breeds and regions associated to egg production in comparison between commercial layers and non-commercial breeds in **chapter 2**, as well as a region harboring a gene associated to dark brown phenotype in the only breed that had dark brown feather in **chapter 3**. Therfore a combination of two methods (F_{ST} and heterozygosity) was used for selection signature detection in **chapter 4**.

Combination of methods

Discovery of the same selected regions using different approaches (ideally approaches with different statistics) can provide convincing evidence of selection signature detection. Several studies have followed this idea for selection signature detection. Grossman *et al.* (2010) proposed a composite of multiple signals (CMS) (i.e. a composite statistic test of several selection signature signals), while Qanbari *et al.* (2014) used two methods with different statistical background (i.e. site frequency spectrum based selection signature detection and linkage disequilibrium based association mapping). However, use of CMS is still challenging in livestock genomic data, because employing calibrated demographic models is required to mimic the empirical data (Qanbari and Simianer, 2014). In this study in **chapter 4,** combination of two methods with different statistic background, i.e. single site differentiation (F_{ST}) and reduced local variability (H_E), was used for selection signature detection. In order to consider a region as under selection, the size of the region was required to be larger than 100 kbp, and also both methods should have been able to detect that region. As demonstrated in Table 2, there is a significant reduction in detection rate when the two methods are combined, compare to using each method separately. If the combination of the two methods was not used, ten-thousands of polymorphisms would have been detected as being under selection.

Table 2. Regions detected as under selection with F_{ST}, H_E and combination of two methods in different contrast.

Contrast	**F_{ST}**	**H_E**	**Combined**
Egg color	228	96	3
Skin Color	231	101	2
Toe number	201	95	3

Additionally, using intra-population methods (i.e. H_E) allowed to identify the population that has been the target of selection. This capability was employed to re-evaluate the regions that have been detected as regions under selection. For instance, the yellow skin chickens were confirmed as the group under selection for the yellow skin color gene, or the brown egg layers chickens were confirmed as the group under selection for the region harboring three QTL associated to brown egg shell color.

Conclusion

Based on the result of this study, it can be concluded that use of inter-population statistics backing with establishment of clear contrast is capable of identifying putative regions harboring genes, QTL and pathways associated to phenotypic variation in the contrast. Using inter-population statistic in this study has led to detection of regions subjected to selection for different traits such as meat production in dual purpose breeds (**chapter 2** and **3**), egg production in layer breeds (**chapter 2**), feather plumage color in breed with dark brown feather (**chapter 3**), skin color in yellow skin color breeds (**chapter 4**) and egg shell color in chickens with brown egg shell (**chapter 4**). Many of the identified regions seem to play important roles in economically important traits of chicken and are good candidates for further studies. However, several of the detected regions in this study (e.g. two of the regions detected in the egg color contrast in **chapter 4**) were not associated with any genes or QTL related to the trait under study, which could be due to insufficient knowledge about these regions (Eyras et al., 2005). Several statistical methods for selection signature detection were explored in this study and a better understanding of the effect of various statistics in selection signature detection was provided. It is demonstrated that the use of different statistics can enhance the detection of different types of selection signatures (e.g. soft sweeps and recent sweeps). Furthermore it is demonstrated that combination of two methods gives a more reliable evaluation for regions detected under selection. In conclusion, it can be claimed that large numbers of populations combined with a clear contrast has a higher effect on the detection power compared to the extreme high resolution (i.e. sequence data) genome scan.

The results discussed throughout this thesis have allowed a better understanding of the effect of different factors on selection signature detection in the chicken genome, but yet, many challenges remain. Further advances in selection signature detection of chicken, particularly for commercial chicken lines, demand establishment of more precise statistical approaches for evaluation of putative regions under selection (e.g. optimization of “combination of methods” explaind in this thesis), and development of statistical methods with high power of detection in a population with high drift. The results discussed in this study may give a better insight about the effect of selection in a diverse set of chicken breeds, and moreover, it may also be used as additional information on regions (or SNPs) that are associated to production or appearance traits in chicken.

References

Akey, J.M., 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19, 711–722. doi:10.1101/gr.086652.108

Akey, J.M., Zhang, G., Zhang, K., Jin, L., Shriver, M.D., 2002. Interrogating a High-Density SNP Map for Signatures of Natural Selection. Genome Res. 12, 1805–1814. doi:10.1101/gr.631202

Amato, R., Pinelli, M., Monticelli, A., Marino, D., Miele, G., Cocozza, S., 2009. Genome-Wide Scan for Signatures of Human Population Differentiation and Their Relationship with Natural Selection, Functional Pathways and Diseases. PLoS ONE 4, e7927. doi:10.1371/journal.pone.0007927

Barendse, W., Harrison, B.E., Bunch, R.J., Thomas, M.B., Turner, L.B., 2009. Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10, 178. doi:10.1186/1471-2164-10-178

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., SanCristobal, M., 2010. Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. Genetics 186, 241–262. doi:10.1534/genetics.110.117275

Browning, S.R., Weir, B.S., 2010. Population Structure With Localized Haplotype Clusters. Genetics 185, 1337–1344. doi:10.1534/genetics.110.116681

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., Nielsen, R., 2005. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 15, 1496–1502. doi:10.1101/gr.4107905

Crawford, R.D., 1990. Poultry Breeding and Genetics. Elsevier, New York.

Elferink, M.G., Megens, H.-J., Vereijken, A., Hu, X., Crooijmans, R.P.M.A., Groenen, M.A.M., 2012. Signatures of Selection in the Genomes of Commercial and Non-Commercial Chicken Breeds. PLoS ONE 7, e32720. doi:10.1371/journal.pone.0032720

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., Jensen, P., Andersson, L., 2008. Identification of the Yellow Skin Gene Reveals a Hybrid Origin of the Domestic Chicken. PLoS Genet 4, e1000010. doi:10.1371/journal.pgen.1000010

Eyras, E., Reymond, A., Castelo, R., Bye, J.M., Camara, F., Flicek, P., Huckle, E.J., Parra, G., Shteynberg, D.D., Wyss, C., Rogers, J., Antonarakis, S.E., Birney, E., Guigo, R., Brent, M.R., 2005. Gene finding in the chicken genome. BMC Bioinformatics 6, 131. doi:10.1186/1471-2105-6-131

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M., Servin, B., 2013. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. Genetics 193, 929–941. doi:10.1534/genetics.112.147231

Grossman, S.R., Shlyakhter, I., Shylakhter, I., Karlsson, E.K., Byrne, E.H., Morales, S., Frieden, G., Hostetter, E., Angelino, E., Garber, M., Zuk, O., Lander, E.S., Schaffner,

S.F., Sabeti, P.C., 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327, 883–886. doi:10.1126/science.1183863

Gunnarsson, U., Kerje, S., Bed'hom, B., Sahlqvist, A.-S., Ekwall, O., Tixier-Boichard, M., Kämpe, O., Andersson, L., 2011. The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. Pigment Cell Melanoma Res. 24, 268–274. doi:10.1111/j.1755-148X.2011.00825.x

Hamblin, M.T., Casa, A.M., Sun, H., Murray, S.C., Paterson, A.H., Aquadro, C.F., Kresovich, S., 2006. Challenges of Detecting Directional Selection After a Bottleneck: Lessons From Sorghum bicolor. Genetics 173, 953–964. doi:10.1534/genetics.105.054312

Hutt, F.B., 1949. II. Variation in eggs, in: Genetics of the Fowl. McGRAW-HILL book company, INC., New York.

Innan, H., Kim, Y., 2008. Detecting Local Adaptation Using the Joint Sampling of Polymorphism Data in the Parental and Derived Populations. Genetics 179, 1713–1720. doi:10.1534/genetics.108.086835

Johansson, A.M., Pettersson, M.E., Siegel, P.B., Carlborg, Ö., 2010. Genome-Wide Effects of Long-Term Divergent Selection. PLoS Genet 6, e1001188. doi:10.1371/journal.pgen.1001188

Kelley, J.L., Madeoy, J., Calhoun, J.C., Swanson, W., Akey, J.M., 2006. Genomic signatures of positive selection in humans and the limits of outlier approaches. Genome Res. 16, 980–989. doi:10.1101/gr.5157306

Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P., Hocking, P.M., Fife, M., Salmon, N., Fulton, J., Strom, T.M., Haberer, G., Weigend, S., Preisinger, R., Gholami, M., Qanbari, S., Simianer, H., Watson, K.A., Woolliams, J.A., Burt, D.W., 2013. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics 14, 59. doi:10.1186/1471-2164-14-59

Narum, S.R., Hess, J.E., 2011. Comparison of FST outlier tests for SNP loci under selection. Mol. Ecol. Resour. 11, 184–194. doi:10.1111/j.1755-0998.2011.02987.x

Nielsen, R., 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154, 931–942.

Nielsen, R., Slatkin, M., 2013. An introduction to population genetics: theory and applications. Sinauer Associates, Sunderland, Mass.

Nuzhdin, S.V., Turner, T.L., 2013. Promises and limitations of hitchhiking mapping. Curr. Opin. Genet. Dev. 23, 694–699. doi:10.1016/j.gde.2013.10.002

Porto-Neto, L.R., Lee, S.H., Lee, H.K., Gondro, C., 2013. Detection of signatures of selection using Fst. Methods Mol. Biol. Clifton NJ 1019, 423–436. doi:10.1007/978-1-62703-447-0_19

Price, A.L., Zaitlen, N.A., Reich, D., Patterson, N., 2010. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11, 459–463. doi:10.1038/nrg2813

Qanbari, S., Hansen, M., Weigend, S., Preisinger, R., Simianer, H., 2010. Linkage disequilibrium reveals different demographic history in egg laying chickens. BMC Genet. 11, 103. doi:10.1186/1471-2156-11-103

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T.M., Fries, R., Nielsen, R., Simianer, H., 2014. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. PLoS Genet 10, e1004148. doi:10.1371/journal.pgen.1004148

Qanbari, S., Simianer, H., 2014. Mapping signatures of positive selection in the genome of livestock. Livest. Sci. 166, 133–143. doi:10.1016/j.livsci.2014.05.003

Qanbari, S., Strom, T.M., Haberer, G., Weigend, S., Gheyas, A.A., Turner, F., Burt, D.W., Preisinger, R., Gianola, D., Simianer, H., 2012. A High Resolution Genome-Wide Scan for Significant Selective Sweeps: An Application to Pooled Sequence Data in Laying Chickens. PLoS ONE 7, e49525. doi:10.1371/journal.pone.0049525

Quilez, J., Short, A.D., Martínez, V., Kennedy, L.J., Ollier, W., Sanchez, A., Altet, L., Francino, O., 2011. A selective sweep of >8 Mb on chromosome 26 in the Boxer genome. BMC Genomics 12, 339. doi:10.1186/1471-2164-12-339

Rao, Y.-S., 2007. Comparative study of SNP diversity and calculation of the effective size of population in chicken. HEREDITAS 29, 1083. doi:10.1360/yc-007-1083

Rubin, C.-J., Megens, H.-J., Barrio, A.M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A.M., Andersson, L., 2012. Strong signatures of selection in the domestic pig genome. Proc. Natl. Acad. Sci. 109, 19529–19536. doi:10.1073/pnas.1217149109

Rubin, C.-J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, Ö., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., Andersson, L., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591. doi:10.1038/nature08832

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. doi:10.1038/nature06250

Teshima, K.M., Coop, G., Przeworski, M., 2006. How reliable are empirical genomic scans for selective sweeps? Genome Res. 16, 702–712. doi:10.1101/gr.5105206

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppälä, E.H., Hansen, M.S.T., Lawley, C.T., Karlsson, E.K., Bannasch, D., Vilà, C., Lohi, H., Galibert, F., Fredholm, M., Häggström, J., Hedhammar, Å., André, C., Lindblad-Toh, K., Hitte, C., Webster, M.T., The LUPA Consortium, 2011. Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. PLoS Genet 7, e1002316. doi:10.1371/journal.pgen.1002316

Wakeley, J., Nielsen, R., Liu-Cordero, S.N., Ardlie, K., 2001. The discovery of single-nucleotide polymorphisms--and inferences about human demographic history. Am. J. Hum. Genet. 69, 1332–1347. doi:10.1086/324521

Yang, S., Li, X., Li, K., Fan, B., Tang, Z., 2014. A genome-wide scan for signatures of selection in Chinese indigenous and commercial pig breeds. BMC Genet. 15, 7. doi:10.1186/1471-2156-15-7

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., Zheng, H., Liu, T., He, W., Li, K., Luo, R., Nie, X., Wu, H., Zhao, M., Cao, H., Zou, J., Shan, Y., Li, S., Yang, Q., Asan, Ni, P., Tian, G., Xu, J., Liu, X., Jiang, T., Wu, R., Zhou, G., Tang, M., Qin, J., Wang, T., Feng, S., Li, G., Huasang, Luosang, J., Wang, W., Chen, F., Wang, Y., Zheng, X., Li, Z., Bianba, Z., Yang, G., Wang, X., Tang, S., Gao, G., Chen, Y., Luo, Z., Gusang, L., Cao, Z., Zhang, Q., Ouyang, W., Ren, X., Liang, H., Zheng, H., Huang, Y., Li, J., Bolund, L., Kristiansen, K., Li, Y., Zhang, Y., Zhang, X., Li, R., Li, S., Yang, H., Nielsen, R., Wang, J., Wang, J., 2010. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science 329, 75–78. doi:10.1126/science.1190371

Zhang, G., Muglia, L.J., Chakraborty, R., Akey, J.M., Williams, S.M., 2013. Signatures of natural selection on genetic variants affecting complex human traits. Appl. Transl. Genomics, SPECIAL ISSUE ON PHARMACOGENOMICS & PERSONALIZED MEDICINE AND SPECIAL ISSUE ON EVOLUTIONARY GENOMICS 2, 78–94. doi:10.1016/j.atg.2013.10.002

Zhou, H., Alexander, D., Lange, K., 2011. A quasi-Newton acceleration for high-dimensional optimization algorithms. Stat. Comput. 21, 261–273. doi:10.1007/s11222-009-9166-3

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Prof. Dr. Henner Simianer, for his guidance, and support throughout my PhD studies. I am grateful to him for giving me the opportunity of working on this exciting topic. I have gained priceless knowledge and skills from him.

I am grateful to Dr. Steffen Weigend for accepting to co-reference this thesis. I am thankful for his friendly supervision through my entire Ph.D. study.

I would like to thank Prof. Dr. Christoph Knorr for being my third supervisor.

My research was made possible through the financial support from the German Federal Ministry of Education and Research (BMBF) within the AgroClustEr “Synbreed – Synergistic plant and animal breeding” (Funding ID: 0315528C).

My sincere thanks also goes to Dr. Bertrand Servin for granting me the opportunity to visit his group, and I thank him and María Inés for sharing their ideas with me, and guiding me through the novel research projects in their lab.

I would like to thank all my colleagues from the Department of Animal Sciences, especially the Animal Breeding and Genetics Group, and in highlights, Christan G., Christan R., Guiyan, Malena, Reza, Saber and Swetlana for cultivating a great scientific and social working environment. Exclusive thanks to Ms. Döring whom helped me with all the work-related and personal matters of life in Goettingen.

Thanks to Silvia for her great friendship and all the pleasant times inside and outside Germany during my stay in Goettingen. Our tea-meeting will remain as a great memory forever.

I would like to thank Arash for supporting me like a big brother from my first day of arrival to Germany.

I would also like to thank my sister, Faezeh, for her comments regarding my English writing, and for motivating me to strive towards my goals.

Lastly, a special thanks to my family. Words cannot express how grateful I am to my mother and father for their never-ending supports through my whole life. My sincere love goes to my parents and siblings for their love and all the moral supports.

Curriculum Vitae

Name:	**Mahmood Gholami**
Date of Birth:	17.04.1985
Place of Birth:	Tehran
Nationality:	Iran

EDUCATION

March 2011-present	**Georg-August-Universität Göttingen** iv. Animal breeding and Genetics v. PhD student	Göttingen, Germany
Oct 2008–Feb 2011	**Justus-Liebig-University of Giessen** vi. Agro-biotechnology master program vii. GPA: 1.9 (A-)	Giessen, Germany
Summer 2008	**Royan Institute** viii. Certification of Molecular Genetics Techniques	Tehran, Iran
2003–2007	**University of Tehran- Faculty of Agriculture** ix. B.Sc. in Agricultural Eng. Animal Science x. GPA of Last 2 Years (60 Units) :15.75/20	Tehran, Iran
2002–2003	**Ershad high school** (Double Majored) xi. Diploma in Mathematics xii. Diploma in Biology	Tehran, Iran
2001–2002	**Bonita Vista High School** xiii. Third grade of high school (grade 11) xiv. GPA: 3.83/4.0	San Diego, CA

AREAS OF INTEREST

xv.	Bioinformatics	xvii.	Biotechnology	xix.	Evolutionary studies
xvi.	Molecular animal Breeding	xviii.	Molecular genetics	xx.	Biostatistics

PUBLICATION

- **Mahmood Gholami**, Christian Reimer, Malena Erbe, Rudolf Preisinger, Annett Weigend, Steffen Weigend, Bertrand Servin and Henner Simianer. **Genome scan for selection in structured layer chicken populations exploiting linkage disequilibrium information.** Submitted to PLOS ONE.

- **Gholami M**, Erbe M, Gärke C, Preisinger R, Weigend A, Weigend S, Simianer S (2014) **Population Genomic Analyses Based on 1 Million SNPs in Commercial Egg Layers.** PLOS ONE 9(4): e94509. doi:10.1371/journal.pone.0094509

- **Gholami M.**, Wubishet A. Bekel, Joerg Schondelmaier and Rod J. Snowdon. **Tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploidy plants.** Plant Biotechnology Journal (2012) 10, pp. 635–645

- **M. Gholami**, M.Siedel, G.Haberer, T.Strom, R. Preisinger, H. Simianer, **"An efficient pipeline for sequence based variant detection in comercial chicken lines"**, Poster, Plant and Animal Genome XX Conference, San Diego, USA, January 14-18, 2012.

- **Mahmood Gholami**, Daniela Zeltner, Renate Schmidt,Wolfgang Ecke, Rod Snowdon**, "Multiplexed 454 amplicon sequencing for high-throughput SNP discovery in Brassica napus candidate genes for oil content"**, Poster, Genomics-based breeding, Giessen, Germany, October 26-28, 2010.

- **Mahmood Gholami**, Daniela Zeltner, Renate Schmidt,Wolfgang Ecke, Rod Snowdon, **"Multiplexed 454 amplicon sequencing for high-throughput SNP discovery in Brassica napus candidate genes for oil content"**, Poster, Brassica 2010/17th Crucifer Genetics Workshop, Saskatoon, Canada, September 5-9, 2010.

RESEARCH AND WORK EXPERIENCE

2013,March-May **Visiting Scientist** at INRA laboratory of Cellular Genetics

- Selection signatures detection using hierarchically structures and haplotype information in chicken

2011,March-May **Visiting Scientist** at Helmholtz Institute for Bioinformatics and Systems Biology

- Bioinformatics training (sequence data analysis)

2010,Sep-Nov **Internship** in Max-Planck institute of plant breeding

- Identification and molecular characterization of DNA damage repair genes in *Arabidopsis thaliana* using natural variation

2010,Jan-2011,Feb **Master thesis** in Giessen university, faculty of plant breeding

- Identification of possible rare SNPs in oil content genes of rapeseed by new 454 sequencing technology (Amplicon Sequencing)

2010, March-July **Research assistant** in Giessen university, faculty of plant breeding

- SSR marker analysis in rapeseed for the verticilium resistance

2009, July–Nov **Research assistant** in Giessen university, faculty of plant breeding

xxi. SSR and AFLP marker analysis in rapeseed for the verticilium resistance.

2009, Winter **Research assistant** in Giessen university, faculty of plant pathology

xxii. Mainly involved in DNA and RNA extraction and real time PCR in two different PhD thesis

2008, Summer **Internship** at Royan Institute

xxiii. Involved in the transgenic goat creation project

2007, Summer **Internship** at the Association of Rodehen Broiler Farmer, Tehran, Iran

xxiv. Control on broiler nutrition, health and feed conversion

LANGUAGE SKILLS

Persian (Farsi)	**English**	**German**
Native speaker	Fluent	Intermediate

HONORS & ME

2003 Accepted to Ranked 1st Animal Science Program in Iran

2003–2004 A member of the Animal Science Student Council, University of Tehran

PROFESSIONAL SKILLS & COMPUTER ENVIRONMENT FAMILIARITIES

Laboratory skill: DNA & RNA extraction, polyacrylamide and electrophoresis gel preparation, primer designing, PCR (QPCR, RT-PCR, TD-PCR), Cloning, Gene transferring (Agro bacterium & gene gun), plant tissue culturing.

Bioinformatics software: GATK, Picard, BWA, MAQ, Samtools, Stampy, Velvet, SOAPdenovo,

Sequence analyzing software: CLC Genomics workbench, DNAStar (GATC), IGVBrowser (Galaxy)

Genetic Software: plabQTl, map maker, CLC sequence viewer.

Feed Formulation Software: Uffda, Spartan, Takeaway Chip Layout and Testing using IC Station

Programming and simulation: R console, Python, BASH, Pascal, Cygwin (Linux-like environment).

Publishing, presentation and image editing: Openoffice**,** Adobe Photoshop, Adobe Acrobat, and Microsoft Office: Word, Excel, and PowerPoint.

Operating systems: Linux/Unix/Ubuntu, Windows Vista/NT/2000/XP, Windows 98, and DOS.

REFERENCES

Available upon request.