PLOS ONE

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], 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

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 [1]. 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 [4] chicken may have been domesticated for cultural purposes such as religion, decoration, and cock fighting rather than for food production [5]. Strong selection of production traits started in the 20th century when commercial breeds were selected for either egg-laying or meat production [6].

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 [7]. 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 [8]. 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 [9] 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} [10]. Genes responsible for phenotypic differences between populations are expected to show large allele frequency differences [11].

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 [12]. 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. [13] 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. [14] 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. [15] studied selective sweeps using the same method described by Rubin et al. [13] 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 [16]. 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.

DNA was isolated using a phenol/chloroform method for the DNA isolation [17] 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 [18]. 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/) [19].

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 [20,21] and maximum likelihood estimation of individual ancestries using ADMIXTURE software with several null hypotheses [22].

F_{ST} calculation and permutation test

To identify the regions under selection, Wright's F_{ST} [9] 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).

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 [23].

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

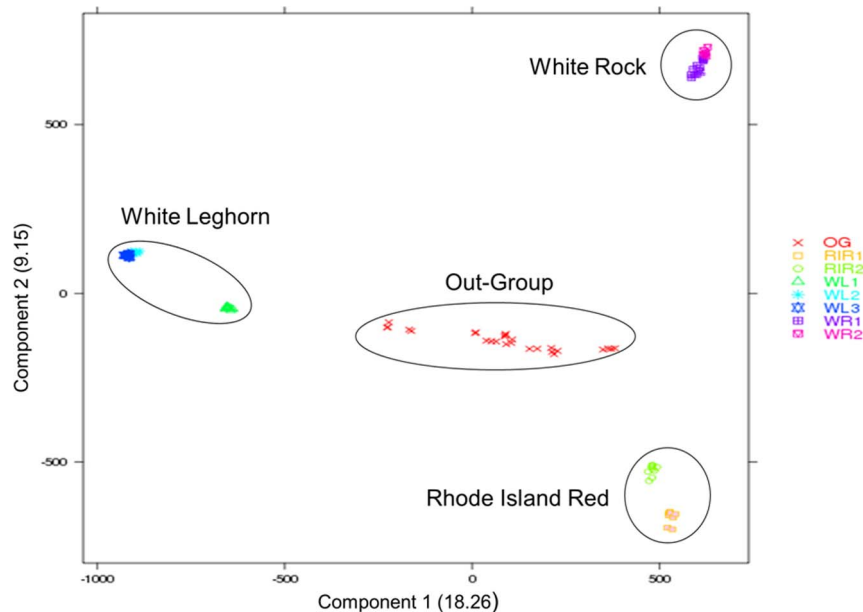doi:10.1371/journal.pone.0094509.t001

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 [24] based on Ensembl data base [25]. For pathway annotation KEGG database [26] 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) [27,28]. We assumed pathways and gene ontologies with p≤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,

Figure 1. PCA analysis for all the 96 individuals with 1 million SNPs.
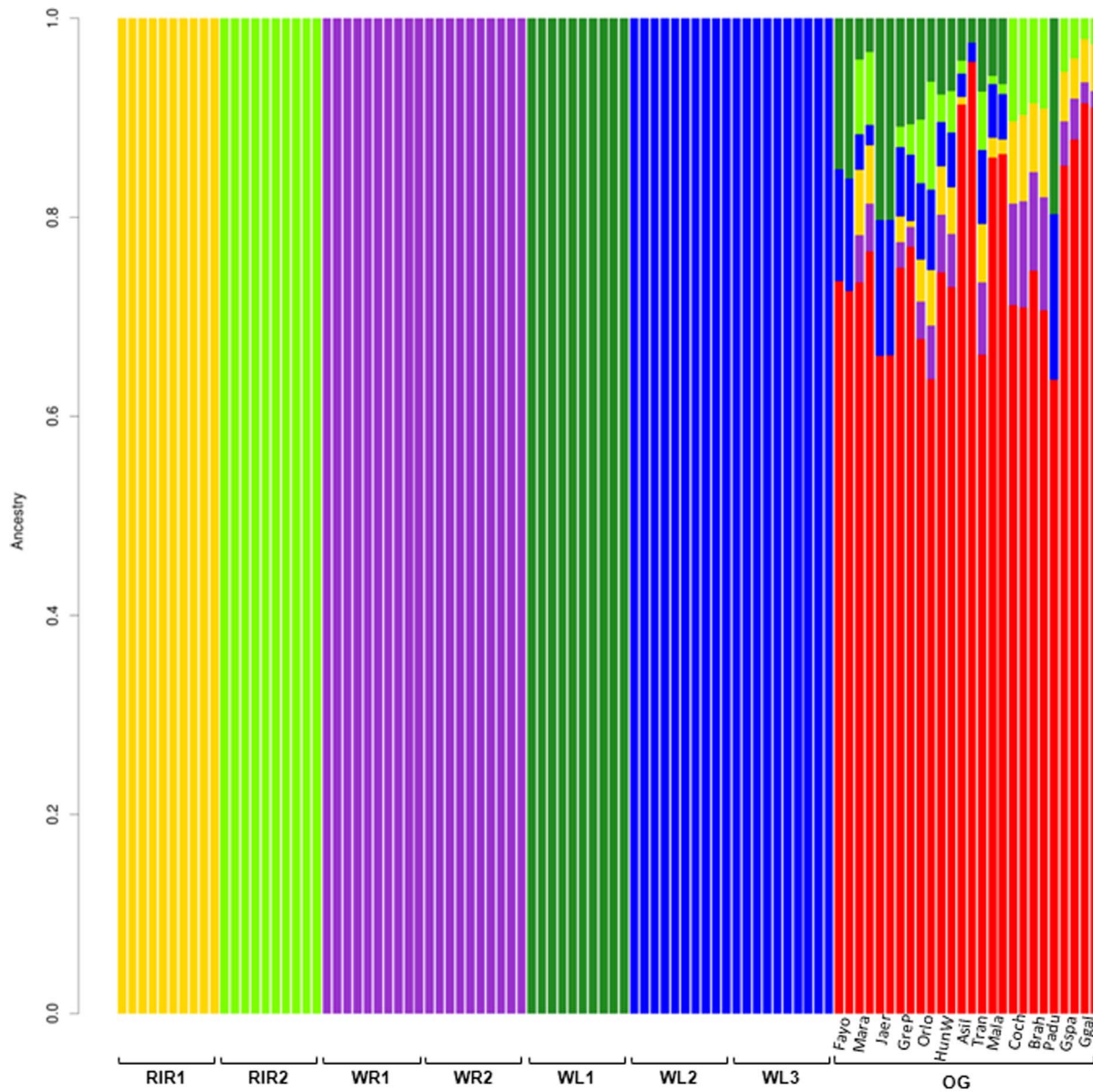doi:10.1371/journal.pone.0094509.g001

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*.
doi:10.1371/journal.pone.0094509.g002

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 [5] 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}

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

	WL	WR
RIR	0.2419($\pm$0.25)	0.1768($\pm$0.20)
WR	0.2641($\pm$0.27)	

values between lines of breeds are always lower than between breeds, which show the similarity within breeds is much higher than between breeds.

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.

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.* [29] 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 [13,30] 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 [31]. However, as Akey [32] 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 [33].

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 [34] (Table 4).

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 [35]. *TGFb2* is expressed 4-fold greater in broiler compared with layer hens at 15 weeks of age [36]. *CAPN1* is associated with meat quality [37,38]. *IGF2* (insulin-like growth factor 2), which is believed to be a major fetal growth factor in contrast to insulin-like growth factor 1 [39], has a great influence on growth and carcass traits in chicken [40]. 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 [5]. *SREBF2* and *RALGAPA1*, which are both associated with reproductive traits and broodiness [41,42], were also located in the regions with high contrast between the two

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

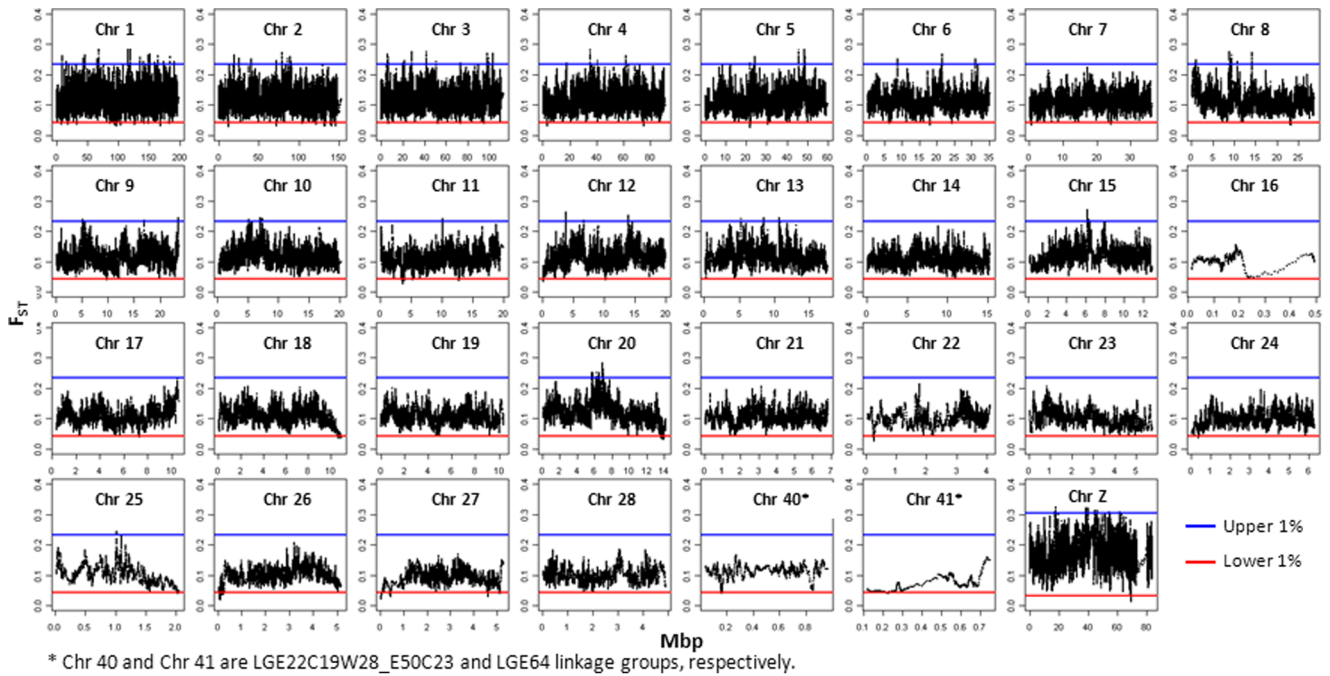|	WL2and3	RIR1	RIR2	WR	OG
WL1	0.1543($\pm$0.21)	0.2653($\pm$0.31)	0.2524($\pm$0.30)	0.2382($\pm$0.29)	0.1184($\pm$0.14)
WL2and3		0.2715($\pm$0.32)	0.2590($\pm$0.30)	0.2567($\pm$0.30)	0.1570($\pm$0.17)
RIR1			0.1148($\pm$0.17)	0.1662($\pm$0.23)	0.1006($\pm$0.13)
RIR2				0.1523($\pm$0.24)	0.0904($\pm$0.11)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

Figure 3. FST-values of overlapping windows for comparison between commercial layers and out-group. Red (blue) line indicates the upper (lower) 1% of FST distribution.
doi:10.1371/journal.pone.0094509.g003

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 [43], along with *SPP1*, which is associated with 5-week body weight and quality of egg shells in laying hens [44],

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
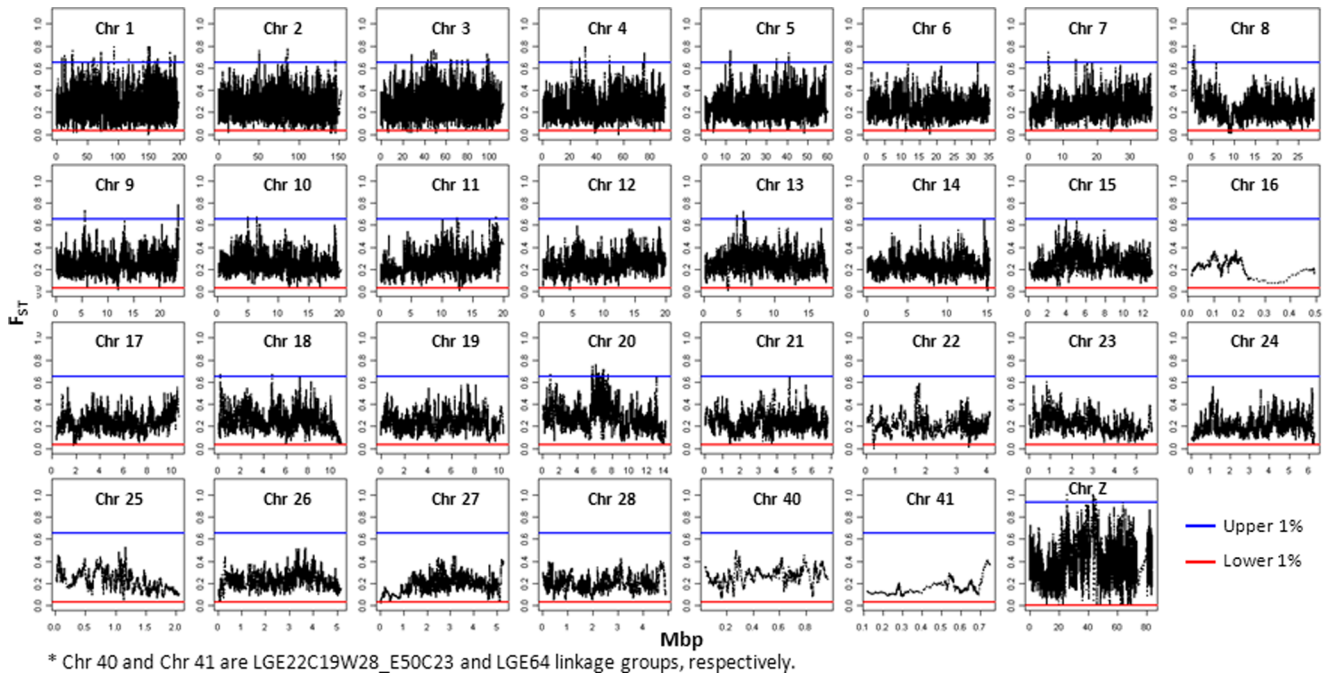

Figure 4. FST-values of overlapping windows for comparison between brown layers and white layers. Red (blue line) indicates the upper (lower) 1% of FST distribution.
doi:10.1371/journal.pone.0094509.g004

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.

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

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.
doi:10.1371/journal.pone.0094509.t004

the regions that differ between commercial-layers and out-group chicken [35,45,46]. 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* [47], is associated with body composition, body weight and affects fatness traits in chickens [48,49], 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 [50], and *IL19* plays an important role in responses to intracellular poultry pathogens like bacteria and protozoa [51]. *POU1F1* and *AMH*, which are both genes related to the growth performance in broiler chickens [52–54], 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.* [13] and Elferink *et al.* [15]. 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 [55]. Although, we have not annotated selection signatures reported in other studies [13,15], 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 [13,15]. We also identified *POU1F1* which binds to and transactivates promoters of growth hormone (GH) and thyroid-stimulating hormone chain (*TSHB*)-encoding genes [56], which were identified by Rubin and Elferink [13,15]. Another reason that the

detected genes are different from the previous studies [13,15] 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

Table S1 List of genes in lower 1% distribution of the comparision of commercial layers and out-group.
(PDF)

Table S2 List of genes in of upper 1% FST distribution in comparision of commercial layers and out-group.
(PDF)

Table S3 List of genes in lower 1% FST distribution in comparision of brown layers and white layers.
(PDF)

Table S4 List of genes in upper 1% FST distribution in comparision of brown layers and white layers.
(PDF)

Table S5 List of pathways and gene ontologies in lower 1% distribution of the comparison of commercial layers and out-group.
(PDF)

Table S6 List of pathways and gene ontologies in upper 1% distribution of the comparison of commercial layers and out-group.
(PDF)

Table S7 List of pathways and gene ontologies in lower 1% FST distribution in comparison of brown layers and white layers. (PDF)

Table S8 List of pathways and gene ontologies in upper1% FST distribution in comparison of brown layers and white layers. (PDF)

References

1. Darwin C (1883) The Variation of Animals and Plants Under Domestication. Macmillan publishers limited.
2. 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.
3. 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.
4. 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.
5. Crawford R (1990) poultry breeding and genetics. Elsevier science.
6. Burt DW (2005) Chicken genome: Current status and future opportunities. Genome Res 15: 1692–1698. doi:10.1101/gr.4141805.
7. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23: 23–35. doi:10.1017/S0016672300014634.
8. 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.
9. Wright S (1949) The Genetical Structure of Populations. Ann Eugen 15: 323–354. doi:10.1111/j.1469-1809.1949.tb02451.x.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. Green MR, Sambrook J (2012) Molecular cloning: a laboratory manual. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
18. Axiom Genotyping Solution Data Analysis Guide (2011).
19. 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.
20. 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.
21. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405. doi:10.1093/bioinformatics/btn129.
22. 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.
23. 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.
24. Durinck S (n.d.) biomaRt: Interface to BioMart databases (e.g. Ensembl, COSMIC,Wormbase and Gramene). R package version 260.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. Muir WM, Aggrey SE (2003) Poultry Genetics, Breeding and Biotechnology. CABI. 724 p.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. 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.
48. 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

Author Contributions

Conceived and designed the experiments: MG ME CG HS. Analyzed the data: MG. Contributed reagents/materials/analysis tools: RP AW SW. Drafted the manuscript: MG ME HS. Read and approved the final manuscript: MG ME CG RP AW SW HS.

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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. 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.