Livestock Science 171 (2015) 1–11

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Prior genetic architecture impacting genomic regions under selection: An example using genomic selection in two poultry breeds

Xinyue Zhang [a,*], Ignacy Misztal [a], Marzieh Heidaritabar [b], John W.M. Bastiaansen [b], Randy Borg [c], Ron Okimoto [c], Robyn L. Sapp [c], Terry Wing [c], Rachel R. Hawken [c], Daniela A.L. Lourenco [a], Zulma G. Vitezica [d], Hans H. Cheng [e], William M. Muir [f]

[a] Department of Animal and Dairy Sciences, University of Georgia, Athens, GA 30605, USA
[b] Animal Breeding and Genomics Center, Wageningen University, 6709 PG Wageningen, The Netherlands
[c] U.S. 412 Scenic, Cobb-Vantress Inc., Siloam Springs, AR 72761-1030, USA
[d] Département de Sciences Animales, Ecole Nationale Superieure Agronomique de Toulouse, Université de Toulouse, Castanet-Tolosan, France
[e] Avian Disease and Oncology Laboratory, 4279 E. MT Hope RD, East Lansing, MI 48823-5338, USA
[f] Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151, USA

ARTICLE INFO

Article history:
Received 8 July 2014
Received in revised form
29 September 2014
Accepted 4 November 2014

Keywords:
SNP
Allele frequency change
Genomic evaluation

ABSTRACT

Background: The objective of this study is to investigate if selection on similar traits in different populations progress from selection on similar genes. With the aid of high-density genome wide single-nucleotide polymorphism (SNP) genotyping, it is possible to directly assess changes in allelic frequencies and regions under selection and address the question. We compared the allele frequencies before and after two generations of selection on an index containing body weight at 6 wk, ultrasound measurement of breast meat, and leg score in two commercial chicken breeds with different selection histories: M breed was primarily selected for rapid growth and commonly used as a broiler breeder sire line; F breed was primarily used as dual-purposed dam line selected for both egg production and growth. Selection was performed on both lines with the same selection intensity and method (Genomic Best Linear Unbiased Prediction, GBLUP, using the single-step approach, ssGBLUP).

Results: After quality control, 52,742 and 52,639 SNPs in M breed and F breed were kept in 4922 and 4904 animals, respectively. The average allele frequency change for both breeds on the autosomes was 0.049. Threshold value for detecting selected regions, where allele frequency changes exceeded expectations under drift were 0.140 and 0.136 for breeds M and F, respectively. According to the criterion used in this study, there were 25 and 17 selection regions detected on breeds M and F, respectively, without any overlap of regions between the breeds. Average heterozygosity change in F breed was greater compared to M breed (0.008 vs. 0.002, $P < 0.01$). Also, there was no overlapping of selected regions with high heterozygosity change between breeds M and F.

* Corresponding author.
E-mail addresses: xinyue@uga.edu (X. Zhang), ignacy@uga.edu (I. Misztal), Marzieh.heidaritabar@wur.nl (M. Heidaritabar), john.bastiaansen@wur.nl (W.M. Bastiaansen), Randy.Borg@cobb-vantress.com (R. Borg), Ron.Okimoto@cobb-vantress.com (R. Okimoto), robyn.sapp@cobb-vantress.com (R.L. Sapp), Terry.Wing@cobb-vantress.com (T. Wing), Rachel.Hawken@cobb-vantress.com (R.R. Hawken), danilino@uga.edu (D.A.L. Lourenco), zulma.vitezica@ensat.fr (Z.G. Vitezica), hans.cheng@ars.usda.gov (H.H. Cheng), bmuir@purdue.edu (W.M. Muir).

http://dx.doi.org/10.1016/j.livsci.2014.11.003

Conclusions: The results indicate that in newly selected populations, even using the same criteria and selection methods, the historical selection goals and breed development determine the loci that most impact selection progress. These results are consistent with quantitative genetic theory that contribution of loci to selection progress depends on initial allele frequency. Therefore it should not be assumed that the same loci will be under selection in different populations even if similar selection goals and methods are used.

1. Introduction

An interesting academic question with practical implications is, "does selection on similar traits in different populations progress from selection on similar genes?". In practice the question is "will genes found to be important in one F breedor a given trait also be important for the same trait in another?". Because selection on traits changes the allelic frequency of the underlying causative genes (Nielsen, 2005), the interspecific to intraspecific variability between populations or between generations increases (Lewontin and Krakauer, 1973). Directional selection is different from other evolutionary factors that either reduce the ratio of within and between population genetic variability, or have no effect on the genetic variability. Selective sweeps, which are genomic region that have recently become fixed due to the selection of advantageous alleles, reduces the variability in the causative genes and flanking sites. To detect incomplete selective sweeps, it should be possible to utilize genome-wide changes in the allele frequency spectrum over time in populations under selection.

For domestic animals, there were already several studies on the allele frequency spectrum of signature of selection by investigating heterozygosity (Elferink et al., 2012), Wright's fixation index (Fst test) (Moradi et al., 2012), and relative extended haplotype homozygosity (REHH test) (Sabeti et al. 2002). However, those studies used cross-generation data during selection, therefore, their results were impacted by both recent and historical selections. Furthermore, most previous studies only have allele frequency data after completion of selection, leaving the initial and change in allele frequencies unknown.

In order to separate the results caused by historical and new selection, our study used two methods: the straightforward allele frequency change from initial to last generation was used to detected genomic change in a recent selection experiment in broiler (meat-type) chickens; and heterozygosity change in above time cession was used to detect selective sweep. Two selection breeds from different origins, a sire breed (M) historically selected for rapid growth, and a dam breed (F) historically selected for both egg production and growth, were used. These breeds were selected for body weight at 6 wk (BW), ultrasound measurement of breast meat (BM), and leg score (LS) using the same index in both. Genotypes on animals in these breeds were collected for genomic selection. From this data, we attempted to identify the changes in allele frequency spectrum across chromosomes for each generation that should provide insights into how the genome responds to selection.

2. Results

2.1. Effect of selection traits on the change of genetic variation

Changes in allele frequency between *G*0 and *G*2 (d_{02}) in M breed and F breed were calculated to compare the response to selection. Whole-genome patterns of allele frequency change in M breed and F breed were different with respect to the positions, the ranges of putative select regions, and values of the most extreme d_{02} (Figs. 1 and 2). Thresholds for significant d_{02} determined by gene dropping method were 0.140 for M breed and 0.136 for F breed (Fig. S1). None of the selected regions were overlapping between the two breeds. The average changes in allele frequency ($\overline{d_{02}}$) on autosomes were the same, 0.049, in both breeds (Table 3). As expected for the sex chromosome, and aggravated by the smaller number of male vs. female parents, chromosome Z had a larger average allele frequency change compared to the autosomes. This change was greater in M breed than that found for F breed (0.070 vs. 0.061, respectively, $P<0.01$). The $\overline{d_{02}}$ of all chromosomes for *M* breed and F breed are 0.051 and 0.049. Also, the average minor allele frequency (MAF) of *G*2 is higher than the MAF of *G*0 in both breeds, again in both the Z chromosome and in autosomes (average MAF difference, autosomes: 0.002 for both breeds; chromosome Z: 0.016 and 0.008 for M breed and F, respectively, $P<0.01$). In selected regions, the average allele frequency changes were 0.177 for M breed, and slightly smaller, 0.176 for F breed, but not significantly different between the breeds ($P=0.7$). The distribution of d_{02} values showed a longer tail in M breed than F breed, indicating that SNPs in M breed have more extreme allele frequency changes after two generations of selection (Fig. 3).

2.2. Selected regions

With both GBLUP selected breeds, less than half of the chromosomes contained extreme regions where the running average of d_{02} exceeded the threshold (Figs. 1 and 2, Tables S2–S6). The threshold was exceeded on 12 and 9 chromosomes, and in 25 and 17 regions, in M breed and F breed, respectively. The total length of selected regions was 11,531 kb and 8396 kb; and the average length was 494 kb and 461 kb for M breed and F breed, respectively. No

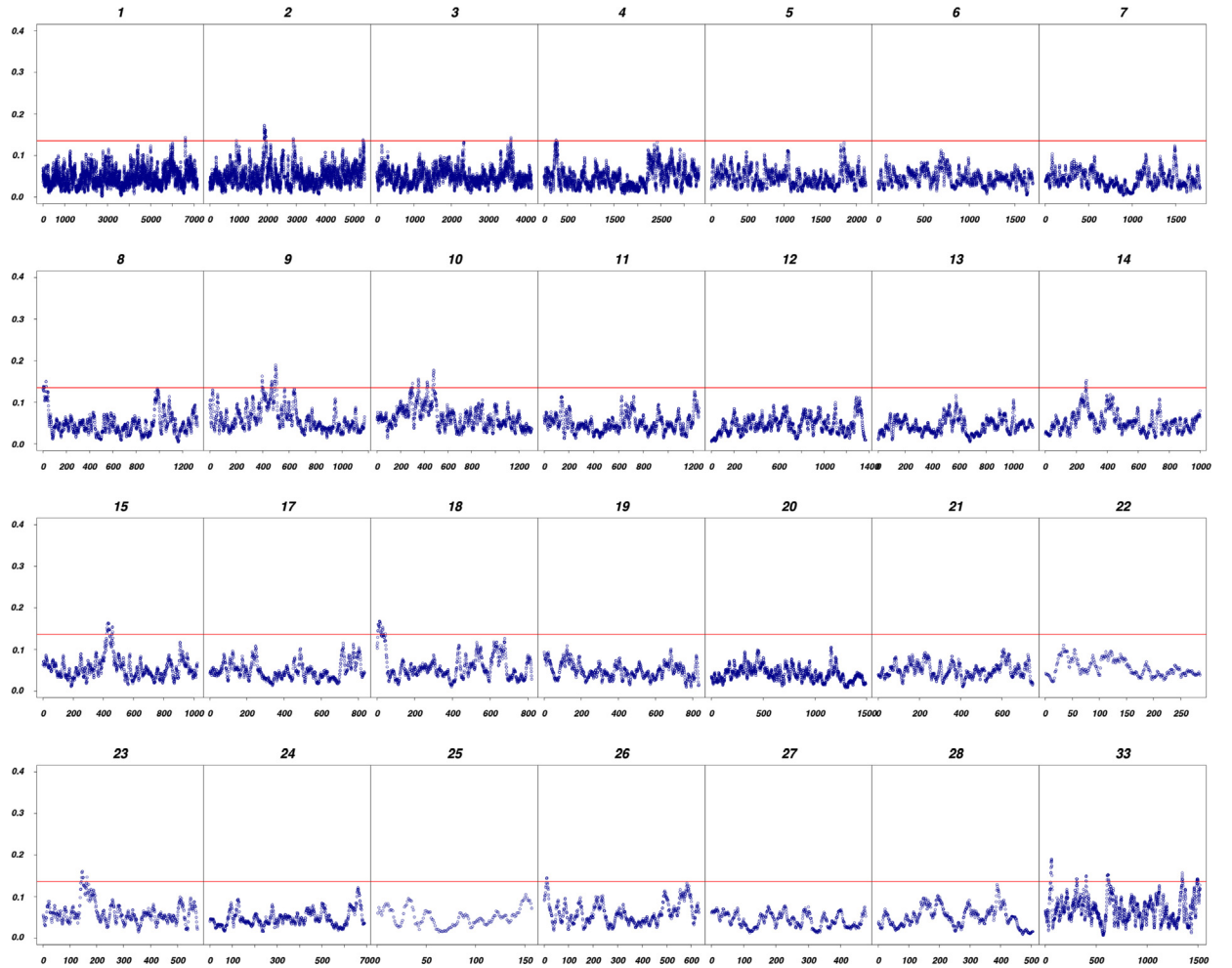

Fig. 1. Pattern of genetic variation after two generations of selection for M breed. Running average of allele frequency distribution of 44,770 SNPs along the whole genome is plotted against the sequence. The deviations above the threshold show signals of selection. Chromosome 33 is chromosome Z.

overlapping regions were found between breeds under resolution of 23 kbp/SNP (Tables S3 and S4). The greatest changes in the running averages of d_{02} values were found on chromosome 2, 9, 10 and Z on M breed; and on chromosomes 4, 12 and Z for F breed (Tables S5 and S6). Total numbers of 322 out of 44,770 and 296 out of 44,895 SNPs surpassed the threshold in breeds M and F, respectively.

2.3. Divergence and genetic variability among the breeds

Heterozygosity was expected to decrease in regions of selection (Allendorf 1986; Barton, 1998; Kim and Stephan, 2002). The results shown in Table 4 indicates that there is a positive average change in heterozygosity ($\overline{H_{02}}$) across all autosomes between *G*0 and *G*2 in both breeds (*G*0–*G*2 = 0.004 for M breed and 0.008 for F breed, $P < 0.01$), and the change in F breed is much larger compare to M breed ($P < 0.01$). However, the Z chromosome has a bigger but increased $\overline{H_{02}}$ (*G*0–*G*2 = −0.086 and −0.088 for M breed and F, respectively, $P < 0.01$).

The threshold values for significant heterozygosity changes (H_{02}) are 0.136 and 0.125 for breeds M and F, respectively. The running average of H_{02} showed multiple regions above the thresholds (Figs. 4 and 5) that overlapped with significantly selected regions based on d_{02} (Figs. 1 and 2) in both breeds. In M breed, chromosomes 2, 3, 9, 10, 15 and 18 each have one region that was identified by both methods. In F breed, one region each on chromosomes 3, 4, 7, 11 and 12, and two regions on chromosome 6 also overlapped between the two methods.

3. Discussion

Our results indicate that both breeds M and F have many genome regions where allele frequency changes are observed after 2 generations of selection. For both breeds, the average MAF was higher in *G*2 than in *G*0, implying a certain level of selection for minor alleles. The average absolute allele frequency changes on autosomes were the same for both breeds, which was expected since they had similar effective size, leading to similar impact of drift, and both were selected for two generations. However, the patterns of d_{02} were vastly different for breeds M and F: non-overlap of selection regions; more and larger selected

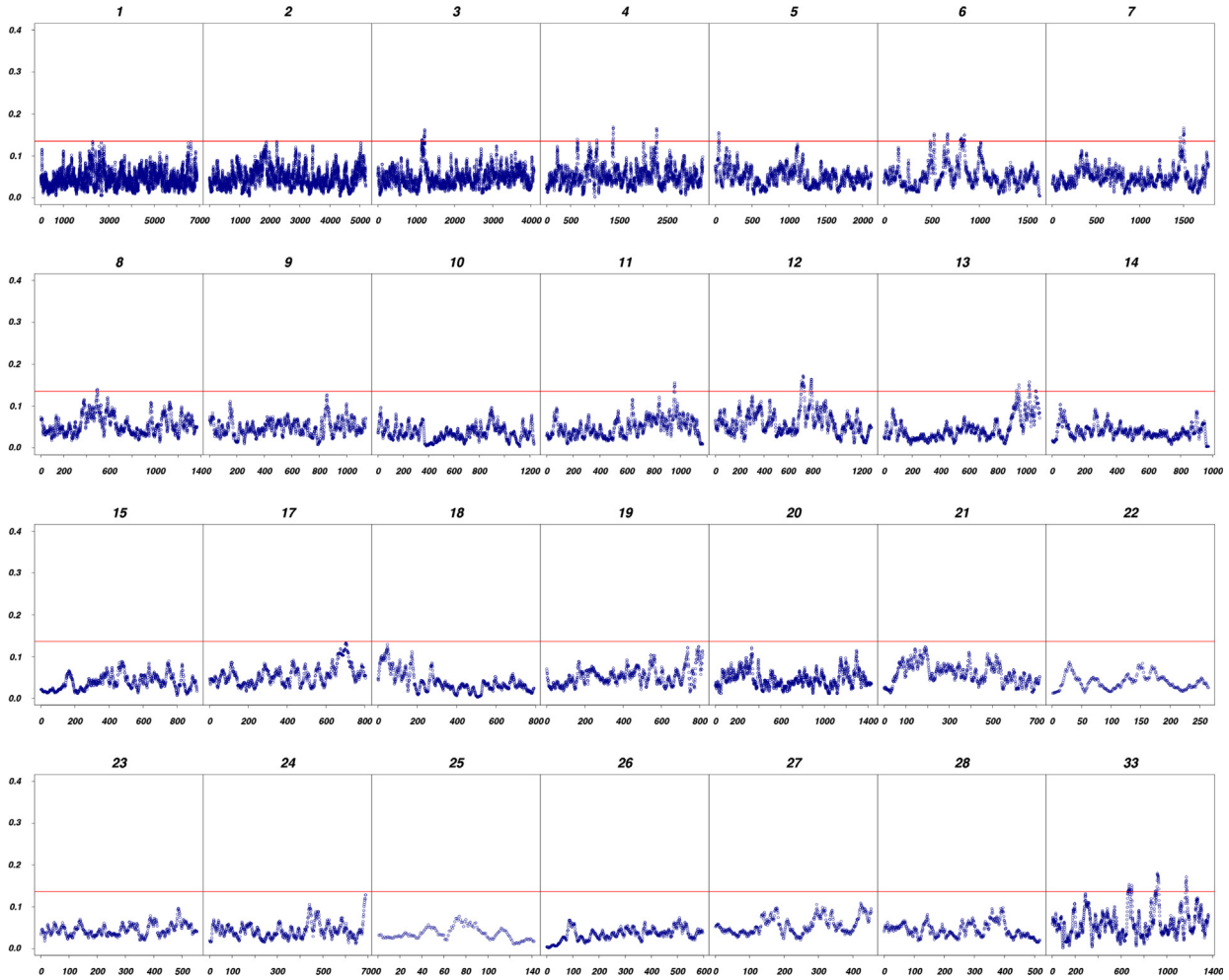

Fig. 2. Pattern of genetic variation after two generations of selection for F breed. Running average of allele frequency distribution of 44,895 SNPs along the whole genome is plotted against the sequence. The deviations above the threshold show signals of selection. Chromosome 33 is chromosome Z.

Table 1
Total number of genotyped animals and number of animals that were selected based on EBV.

Breed	Total genotyped animals	Selected animals			
		G0		*G2*	
		Female	Male	Female	Male
M	4922	200[a]	20	200	200
F	4904	200	20	20	20

[a]Number of selected animals were from Cobb-Vantress Inc.

Table 2
Number of genotyped animals retained after QC.

Breed	*G0*	*G2*	Total
M	1165	1009	4871
F	1154	1028	4774

regions in M breed compared to F breed; and also larger d_{02} values in M breed than in F breed. The larger number of selected regions implies that more genes or functional elements were selected and on top of that the larger peaks in d_{02} indicates a stronger selection on those regions.

The breeds experienced the same recent selection goal and intensity, density of genomic data, and had similar effective population sizes, which means that the distinct genetic backgrounds of M breed and F were responsible for the diversity in their allele frequency changes (Falconer 1960). QTLs associated with the traits in breeds M and F started at different initial allele frequencies, given their differences in genetic architecture due to different historical selection goals, numbers and sizes of QTL affecting the traits, and LD (Lewontin, 1988). The initial allele frequency of biallelic loci determines how and how strong the frequency changes along selection (Kimura, 1957), assuming no selective advantage to consider, as it was not natural selection. The larger selected regions and d_{02} in M breed compared to F breed indicates association with historical selection, where M breed was selected

historically on growth traits with higher heritability than reproductive traits in which selection of F breed was based on. Nevertheless, for F breed there might be more genes but less selection intensity involved in the historical genetic architecture due to dual-purpose selection, resulting in the selection response being more distributed across the genome and fewer regions that pass the threshold. An interesting finding was that regions of d_{02} peaks appeared in breeds M and F were totally different. This attests that the same selection goal does not necessarily mean selection of the same genes, even in the same species. For example, alleles already fixed in M breed would not change in frequency, but still could be selected in F breed.

Table 3
Average difference in allele frequencies ($\overline{d_{02}}$) and major allele frequencies (f) of autosomes and chromosome Z within generations between breeds.

Breed		*G2–G0*		*G0*		*G2*	
		1–28[a] $\overline{d_{02}}$	**Z** $\overline{d_{02}}$	**1–28** f_0[d]	**Z** f_0	**1–28** f_2[e]	**Z** f_2
M	Average[b]	0.049	0.07	0.734	0.809	0.731	0.793
	SD	0.04	0.051	0.142	0.153	0.14	0.149
	N[c]	43,250	1520	44,056	1659	43,542	1533
		$\overline{d_{02}}$	$\overline{d_{02}}$	f_0	f_0	f_2	f_2
F	Average	0.049	0.061	0.739	0.748	0.736	0.74
	SD	0.029	0.048	0.143	0.152	0.142	0.148
	N	43,533	1362	42,629	1414	42,180	1375

[a] Chromosome 16 excluded.
[b] MAF=0 excluded.
[c] *N*: total number of SNPs.
[d] f_0: allele frequency at generation 0.
[e] f_2: allele frequency at generation 2.

Unlike allele frequency change and heterozygosity that could be affected by recent and historical events, genomic selection response based on GEBV changes of SNP in a region, however, measured the change of genetic effect responding to the current selection. Results demonstrate that most putative selection regions based on allele frequency change did not show peak values in genomic selection response (Tables S7 and S8). For autosomes, out of 19 regions in M breed, only 1 region exceeded 3 standard deviations from the mean (region 17, chromosome 23); and no region in F breed exceeded 3 standard deviations from the mean. Also, no region in either breed had top GEBV at *G2* that ranked outside 3 standard deviations from the mean. The low heritability (0.24, 0.27 and 0.12 for BW, BM, and LS, respectively) may explain part of the inconformity. More importantly, based on the above assumption of historical selection, the putative selection regions were more affected by historical selection hence might not necessarily overlap with current selection regions.

Ascertainment bias is not supposed to influence our results, although it was created at the same time the SNP chip was created because only common SNPs were placed on the chip. The impact of ascertainment bias is that rare alleles were neither present nor tested for. However this bias has little or no impact on our results for 2 reasons: (1) if rare alleles were selected for or against, the contribution of those SNPs to the total genetic variation will be small since that contribution is $2p(1\text{-}p)u^2$, where p is MAF and u is marker effect, meaning that even if the effect of the allele is large, the weight will be small. Such loci will eventually contribute to total genetic variation as the allele frequency approaches 0.5. (2) Because we assumed the most conservative setting, i.e. $P=0.5$, if such alleles were present, and could have been tested for, we most likely would not have detect them due to the small effect such rare alleles have on genetic variance.

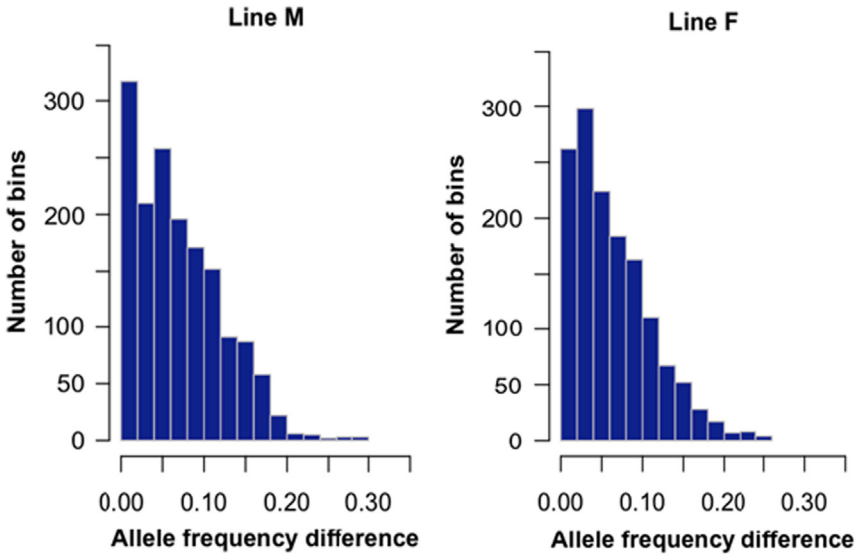

Fig. 3. The distribution of d_{02} after two generations of selection on GBLUP breeding values. *X*-axis is d_{02} value, and *y*-axis is the number of bins.

In other studies, QTL have been discovered across the whole genome, located on all macro-, intermediate-, micro-chromosomes and on chromosome Z. Using the same populations analyzed in this study, Wang (2013) identified the top 10 genome regions that explained genetic variance of the 3 traits that breeds M and F were selected on. These

Table 4
Average heterozygosity (H_p), mean difference of heterozygosity ($\overline{H_{02}}$) and standard deviation by breeds and generations.

Chromosome		M			F		
		G2–G0 $\overline{H_{02}}$	***G0*** H_{p0}[d]	***G2*** H_{p2}[e]	***G2–G0*** $\overline{H_{02}}$	***G0*** H_{p0}	***G2*** H_{p2}
1–28[a]	Average	−0.004	0.362	0.36	−0.008	0.358	0.353
	SD[b]	0.062	0.142	0.145	0.058	0.143	0.148
	N[c]	45,063	44,056	43,542	42,161	42,629	42,180
		$\overline{H_{02}}$	H_{p0}	H_{p2}	$\overline{H_{02}}$	H_{p0}	H_{p2}
Z	Average	0.086	0.027	0.126	0.094	0.051	0.146
	SD	0.056	0.027	0.735	0.051	0.063	0.082
	N	1502	1659	1533	1372	1414	1375

[a] Chromosome 16 excluded.
[b] MAF=0 excluded.
[c] *N*: total number of SNPs.
[d] H_{p0}: heterozygosity at generation 0.
[e] H_{p2}: heterozygosity at generation 2.

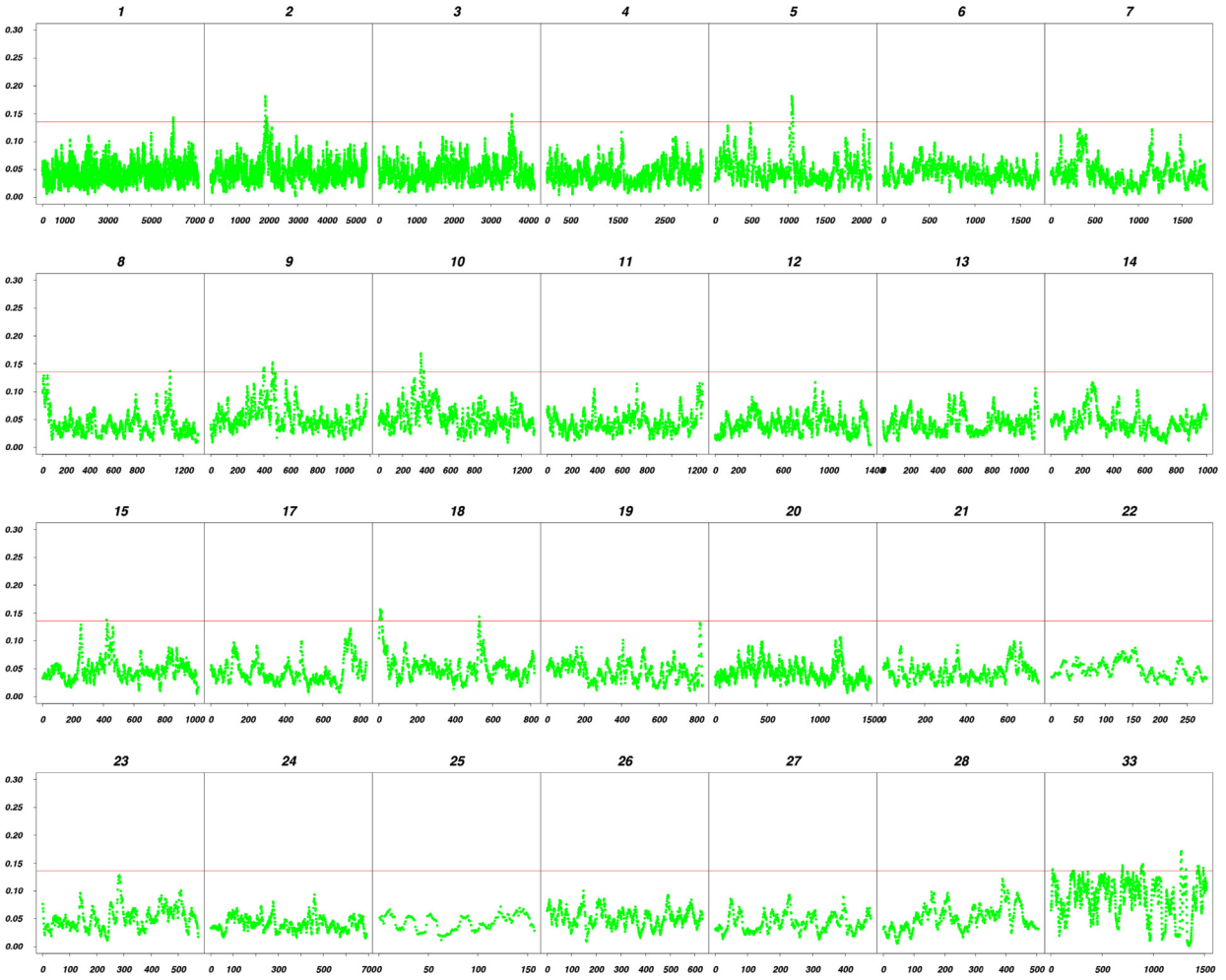

Fig. 4. Pattern of heterozygosity after two generations of selection for M breed. Running average of allele frequency distribution of 46,293 SNPs along the whole genome is plotted against the sequence. The deviations above the threshold show signals of selection. Chromosome 33 is chromosome Z.

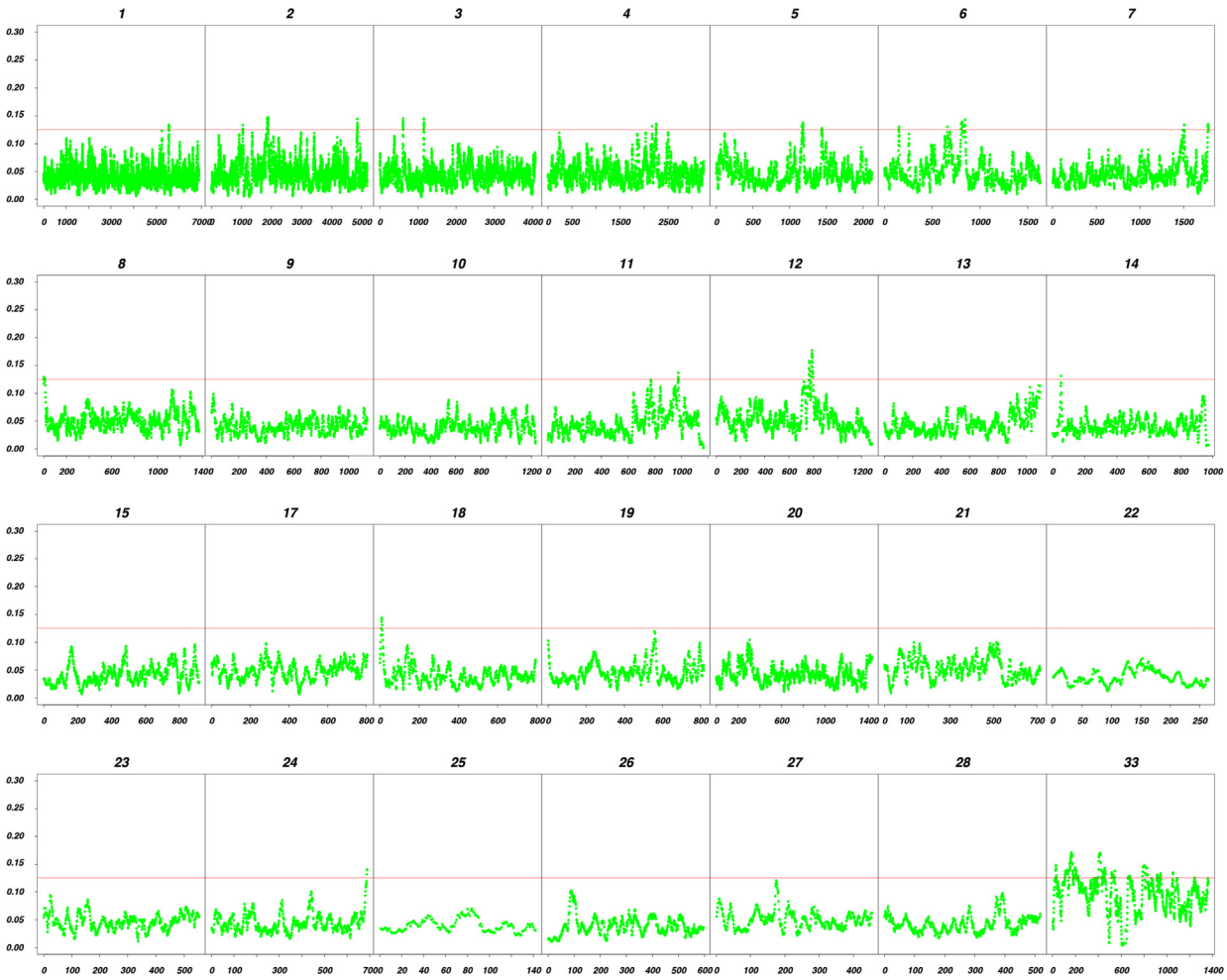

Fig. 5. Pattern of heterozygosity after two generations of selection for F breed. Running average of allele frequency distribution of 43,253 SNPs along the whole genome is plotted against the sequence. The deviations above the threshold show signals of selection. Chromosome 33 is chromosome Z.

associated genome regions were detected using classical GWAS with WOMBAT (Meyer, 2007), ssGBLUP, and Bayes B methods. Only one of the selected regions identified in our study overlaps with the associated regions found by Wang (2013). The overlap was found in F breed, where the selected region is located on chromosome 6, from 19,539,027 bp to 20,308,725 bp. The corresponding region was associated with body weight at 6 wk was located from 19,470,652 bp to 19,901,892 bp and explained 5.97% of genetic variance according to the WOMBAT analysis (Wang 2013). The ssGBLUP and BayesB methods also identified an association in this region between 19,916,663 bp and 20,267,429 bp, accounting for 2.2% and 4.24% of the genetic variance, respectively. The lack of consistency between association and selection results could due to genetic drift, mutation rate, as well as the fact that the model only accounted for additive genetic effect. On the other hand, genetic analysis only seeks for effective SNPs, not their favorable haplotypes. Lastly, the results of association analysis were shown to be method-sensitive (Wang, 2013). Current experience with GWAS indicates that although many associations are detected in several regions, only a few of them are found in similar studies.

Heterozygosity changes on autosomes from *G*0 to *G*2 indicated that selection reduces heterozygosity ($P < 0.05$). The pattern is different from d_{02}, but both breeds have overlap between d_{02} peaks and H_{02} peaks. The overlapped regions confirmed that certain haplotypes have been selected within those areas. The peak regions that only appear in d_{02} but not in H_{02} can occur when a haplotype that was favored contains the minor allele for some SNPs and the major allele for others. The peak regions that only appear in H_{02} but not in d_{02} may indicate no unique haplotype was favored. Heterozygosity pattern of past selection gives the position of selective sweep (Barton, 1998; Kim and Stephan, 2002), which is a wide range of adjacent alleles that became fixed under strong directional selection. In our case, a change of heterozygosity, instead of fixation, was used to identify selective sweeps due to recent selection.

Chromosome Z is different from autosomes in a number of ways, e.g., higher major allele frequencies, larger $\overline{d_{02}}$, lower average heterozygosity but larger $\overline{H_{02}}$ from *G*0 to *G*2, which

increased rather than decreased as generation of selection increases. It is important to note that heterozygosity analysis was done with genotypes of males only, as females are hemizygous (ZW) in chicken. Sundström et al. (2004) observed that when male effective population sizes are smaller, as is the case in many livestock selection programs, a selective sweep will reduce levels of genetic variability on the Z chromosome more drastically than on autosomes. Moreover, the recombination rate on Z chromosome is about 1.3 cM/MB, ~2.5 times less than the average autosomal recombination rate (Levin et al., 1993), thus the effects of selection on linked neutral sites on chromosome Z would stretch much farther on average than autosomes. As expected from these observations, we found that heterozygosity on chromosome Z changed more drastically than on autosomes in both breeds. Interestingly though, the genetic variability raised rather than reduced. Storchova and Divina (2006) found enrichment of male-biased genes (genes expressed preferentially or exclusively in male, e.g. genes coding sperm) but underrepresentation of female-biased genes on chicken Z chromosomes. Bellott et al. (2010) found that chicken Z chromosomes are more uniquely responding to selection for traits that benefit male sex traits more than female. Therefore, the heterozygosity increases on chromosome Z under selection was probably linked to male sex traits indirectly affected by selection breed.

Average heterozygosity of pooled autosomes and sex chromosomes in M breed ranged from 0.346 to 0.352. Elferink et al. (2012) used the same SNP array on commercial and non-commercial chickens where heterozygosities ranged from 0.39 to 0.43 for broiler sire breeds, and 0.35 to 0.42 for broiler dam breeds. These values are larger than in layers, given larger Ne and possibly less historic selection intensity in broilers compared with layers.

Previous studies investigating selective sweeps on domestication of chicken also showed effects of selection on genetic variability (Elferink et al., 2012; Wang, 2013). These studies analyzed the genetic variation across current generations to discover the impact of past selection. Of our 41 putative selected regions, 6 of them overlapped with reference studies (Tables S6), however, most selective sweeps from previous studies do not show overlap with our results, presumably because the fixation in their results was generated by historical selection, and our study of recent selection cannot change allele frequencies in a large scale if those regions were already under fixation process. Previous studies confirmed that selection is the major cause of the frequency spectrum pattern change on chromosomes.

3.1. Characterizing biological functions in putative selected regions

We were also interested to see if QTLs within the selected regions overlap with known QTLs from chicken. QTL were identified from the animal QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/GG/index) and compared with selected regions found in our study. The QTL in regions with large d_{02} changes were found to be related to either production or health. Selected regions in M breed overlapped with 4 QTL for body weight, 2 for growth, 1 for residual feed intake, 1 for muscle weight, 1 for muscle size, 1 for number of eggs, and 1 for age at first egg. Selected regions in F breed overlap with 1 QTL for egg shell thickness, 3 for carcass weight, 2 QTL for carcass components, 1 QTL for feather pecking, and 6 QTL associated with health traits.

Of all the genes located within the selected regions, the interesting candidate genes are listed in Tables S9. Of interest in M breed, carboxypeptidase B1 (CPB1) is located in the highest peak d_{02} region on chromosome 9. Carboxypeptidase B1 is a necessary enzyme especially in the processing of recombinant insulin, and insulin is a vital hormone regulating the carbohydrate and fat metabolism in the body (Ladisch and Kohlmann, 1992). Epidermal growth factor receptor (EGFR-CHICK) located on chromosome 2, which is also the 3rd highest peak of d_{03}, EGF stimulates the growth of various epidermal and epithelial tissues in vivo and in vitro and of some fibroblasts in cell culture (Groenestege et al., 2007). MYOCD (myocardin) is located in the 4th region on chromosome 18 and plays a crucial role in cardiogenesis and differentiation of the smooth muscle cell lineage (myogenesis) (Du et al., 2003). Adenylated cyclase 10 (ADCY10) on chromosome Z has a critical role in mammalian spermatogenesis. In human, it produces the cAMP which mediates in part the cAMP-responsive nuclear factors indispensable for maturation of sperm in the epididymis. It induces sperm capacitation and is involved in ciliary beat regulation. (Geng et al., 2005; Schmid et al., 2007).

In F breed, on chromosome Z, the highest d_{02} region contains lipoprotein lipase (LPL), which catalyze the hydrolysis of triglycerides of circulating chylomicrons and very-low-density lipoproteins (VLDL) (Nilsson-Ehle et al., 1980). In the 4th region on chromosome 4, the gene zygotes arrest 1 (ZAR-1) is found, which in human is essential for female fertility and may play a role in the oocyte-to-embryo transition (Wu et al., 2002). The second highest peak contains lipid phosphate phosphatase-related protein type 1-like motif (LOC427306), but its function is uncharacterized so far.

These genes associated in breeds M and F validate our assumption that different genes respond to the same selection direction in different breeds. Further studies on molecular pathways may need to illustrate the mechanism of different response.

4. Conclusions

The effect of selection goals and breeds on change of genomic variation was investigated across the entire genome of two breeds of broiler chicken. Twenty-five and seventeen regions with evidence of selection were detected after GBLUP selection in a male and a female broiler breeds, respectively. Our study shows that even using GBLUP and the same selection method (GBLUP) and the same selection index, changes in genomic variation are different between breeds. Given that both breeds have the same genes, this result implies that the historical goal during breed development changed the genetic architecture of each breed such that the regions currently selected were altered. These results are consistent with quantitative genetic theory that the contribution of loci to selection progress is dependent on initial allele frequency. Also,

several QTLs overlap with the regions detected by allele frequency and heterozygosity changes indicating that these methods may have potential to identify genes that are functionally linked to the breeds.

5. Methods

5.1. Data structure

Data was provided by Cobb-Vantress Inc. (Siloam Springs, AR). Animals from two pure breeds of commercial broilers were used. M breed was characteristic of a line primarily selected for rapid growth and commonly used as a primary broiler breeder sire line, and F breed was characteristic of a dual-purpose line selected for egg production and growth and commonly used as primary broiler breeder dam breed. In the experiment, both breeds were selected at 6 wk of age for body weight (BW, g), ultrasound measurement of breast meat (BM, cm^2), and leg score (LS, 'acceptable' or 'not acceptable') (Chen et al., 2011). The initial training dataset contained 2000 animals from 2 generations ($G-1$ and $G0$), which was used to estimate SNP effects. From $G0$, selection was performed for 3 generations with about 800 animals genotyped as selection candidates in each generation of each breed. Then about 20 males and 200 females were selected for breeding. ssGBLUP method was used for estimation of genomic breeding values (GEBV) (Aguilar et al., 2011), except for $G-1$ of F breed where selection for LS was done with GEBV from a BayesA (Meuwissen et al., 2001). The initial data set for the prediction of GEBV of animals in generation $G0$ contained 183,784 and 164,246 broilers in M breed and F breed, respectively.

Pedigree information included sires and dams without records in 2 historical generations and 3 selection generations. The total number of records at the end of the experiment was 297,017 for M breed, and 277,051 for F breed.

5.2. Genotype data

Genotypes for 57,636 SNPs were obtained using the chicken IIIumina Infinium iSelect Beadchip (Groenen et al., 2009). Total number of genotyped animals was 4922 in M breed and 4904 in F breed (Table 1). In M breed, 4994 SNPs were removed because the call rate was less than 0.90, the MAF was 0, or the location was on unassigned chromosomes or incomplete chromosomes (16 and W); 51 animals were removed because of low call rate (<0.90) or parent-progeny conflicts. In F breed, 4997 SNPs and 130 animals were similarly removed.

5.3. Breeding structure

The populations spanned several generations. $G0$ was the base population randomly selected from a historical set $G-1$, generation $G1$ were offspring of randomly selected parents from $G0$, $G2$ were offspring of parents selected from $G1$ on the index, and finally $G3$ were offspring of parents selected from $G2$ on the index. Allele frequency differences were obtained between all animals in $G0$ and all animals in $G2$, which are separated by 1 generation of random and 2 generations of directional selection. In other words, $G3$ data represented selected animals in $G2$, whereas $G0$ data represented the same generation.

5.4. Allele frequency changes

Allele frequencies (f) were computed in $G0$ and $G2$ by counting. The absolute values of changes in allele frequencies ($d_{02}=|f_2-f_0|$) between two generations within each breed were calculated. Large allele frequency differences in allele frequencies between $G0$ and $G2$ generations were considered as putative selected regions. The running averages of 11 adjacent d_{02} values were plotted against the location of the middle SNP along chromosomes to emphasize the systematic changes of frequencies in a window. Window size of 11 was chosen on a criterion if the frequency distribution is clear enough but not too disintegrated (results not shown). Threshold values for significant changes in allele frequency were obtained by simulating the flow of alleles through the real pedigree, and gene dropping performed using HaploSim (Coster and Bastiaansen, 2010). Haplotypes were simulated with 20 loci along one chromosome with 0 mutation rate and 0.5 initial allele frequency. A 0.5 starting frequency gives the largest possible drift variance and leads to a conservative threshold. The haplotypes were simulated for the founder animals (1165 animals in M breed, and 1154 in F breed) in the pedigree (Table 2). Genotypes were subsequently assigned to offspring according to Mendelian transmission rules. The changes in allele frequency from $G0$ to $G2$, d_{02}, were computed for 1000 replicates. Then a distribution of the d_{02} was obtained from the 1000 replications of 20 SNPs. A threshold for evidence of selection was determined as the 95% upper bound of the distribution obtained under drift ($P<0.05$).

5.5. Genetic variability between the breeds

Genetic variability was assessed by looking at absolute heterozygosity change between $G0$ and $G2$ (H_{02}) calculated in an overlapping sliding window approach with window size of 5 (Rubin et al., 2010). The equation for heterozygosity within generation is $H_P = 2\sum n_{MAJ}\sum n_{MIN}/(\sum n_{MAJ}+\sum n_{MIN})^2$, where $\sum n_{MAJ}$ is the sum of major allele frequencies, and $\sum n_{MIN}$ is the sum of the MAF in a window. Then the genetic variability was calculated as the absolute difference between H_P of $G0$ and $G2$. The threshold for extreme high or low heterozygosity change was defined as 4 times the standard deviation of genetic variability across chromosomes.

5.6. Genomic selection response from ssGBLUP

GEBV of putative selection regions and genomic selection response was estimated by using the mixed model below:

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z+\lambda * H^{-1} \end{pmatrix}\begin{pmatrix} b \\ u \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix},$$

where X and Z are the incidence matrices corresponding to fixed effects and additive genetic effects, respectively; b is a vector of fixed effects including an overall mean, hatch number and breed; u is the vector of random additive direct genetic effects; λ is the ratio of residual to additive genetic variances; $H-1$ is the inverse of a matrix that combines pedigree and genomic relationships (Aguilar et al., 2010); y is the vector of phenotypic records, in a multi-trait scenario; and e is the vector of residual effect, assuming independent and followed a normal distribution $e \sim (0, I\sigma_e^2)$.

GEBV of SNP regions were calculated as the summation of SNP content times SNP effects for that region for all genotyped animals within a generation. Thereafter, genomic response was accounted for by the change of average GEBV from $G0$ to $G2$. GEBV and genomic selection response for each of the three traits were analyzed separately for each of the two breeds. The sex chromosomes were excluded from the model.

Conflict of interest

The authors declare that they have no competing financial, professional or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

Author's contributions

XZ carried out the study, performed the data analysis and drafted the manuscript. IM participated in the design and coordination of the study. MH participated in the data analysis and helped to draft the manuscript. JWMB participated in the design, provided the software package for data analysis and helped with the data analysis. RRH directed the experiment and provided the data. RO carried out the selection experiment, and advised in the design of the study. RLS provided the data. RB provided the data. TW provided the data. DAL helped with the data analysis. VGZ revised the manuscript. HHC participated in the design and coordination. WMM participated in the design and coordination. All authors read, revised and approved the final manuscript.

Acknowledgments

We thank USDA, United States, National Institute of Food and Agriculture Animal Genome Program for providing the Grant (no. 2009-65205-05665). We also thank Cobb-Vantress Inc. for providing the data and selection experiments.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.livsci.2014.11.003.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., Lawlor, T.J., 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93, 743–752.

Aguilar, I., Misztal, I., Legarra, A., Tsuruta, S., 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. J. Anim. Breed. Genet. 128 (6), 422–428.

Allendorf, F.W., 1986. Genetic drift and the loss of alleles versus heterozygosity. Zoo Biol. 5 (2), 181–190.

Barton, N.H., 1998. The effect of hitch-hiking on neutral genealogies. Genet. Res. 72, 123–133.

Bellott, D.W., Pyntikova, H.S.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 (29), 612–617.

Chen, C.Y., Misztal, I., Aguilar, I., Tsuruta, S., Meuwissen, T.H.E., Aggrey, S.E., Wing, T., Muir, W.M., 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens. J. Anim. Sci. 89, 23–28.

Du, K.L., Ip, H.S., Li, J., Chen, M., Dandre, F., Yu, W., Lu, M.M., Owens, G.K., Parmacek, M.S., 2003. Myocardin is a critical serum response factor cofactor in the transcriptional program regulating smooth muscle cell differentiation. Mol. Cell. Biol. 23, 2425–2437.

Coster, A., Bastiaansen J.W.M., 2010. HaploSim. ⟨http://cran.r-project.org/web/packages/HaploSim/index.html⟩ [R package version 1.8-4].

Elferink, M.G., Megens, H.J., Vereijken, A., Hu, X., Crooijmans, R.P., Groenen, M.A., 2012. Signatures of selection in the genomes of commercial and non-commercial chicken breeds. PLoS One 7 (2), e32720.

Falconer, D.S., 1960. Introduction to Quantitative Genetics. Ronald press, New York.

Geng, W., Wang, Z., Zhang, J., Reed, B.Y., Pak, C.Y.C., Moe, O.W., 2005. Cloning and characterization of the human soluble adenylyl cyclase. Am. J. Physiol. 288, C1305–C1316.

Groenestege, W.M.T., Thebault, S., van der Wijst, J., van den Berg, D., Janssen, R., Tejpar, S., van den Heuvel, L.P., van Cutsem, E., Hoenderop, J.G., Knoers, N.V., Bindels, R.J., 2007. Impaired basolateral sorting of pro-EGF causes isolated recessive renal hypomagnesemia. J. Clin. Investig. 117, 2260–2267.

Groenen, M.A.M., Wahlberg, P., Foglio, M., Cheng, H.H., Muir, H.J.M., Wong, G.K., Gut, I., Anderson, L., 2009. A high density SNP-based linkage map of the chicken gene reveals sequence features correlated with recombination rate. Genome Res. 19, 510–519.

Kim, Y., Stephan, W., 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160, 765–777.

Kimura, M., 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28, 882–901.

Ladisch, M.R., Kohlmann, K.L., 1992. Recombinant human insulin. Biotechnol. Prog. 8 (6), 469–478.

Levin, I., Crittenden, L.B., Dodgson, J.B., 1993. Genetic map of the chicken Z chromosome using random amplified polymorphic DNA (RAPD) markers. Genomics 16, 224–230.

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

Lewontin, R.C., 1988. On measures of gametic disequilibrium. Genetics 120, 489–853.

Meuwissen, T.H.E., Hayes, B., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Meyer, K., 2007. WOMBAT-A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. Sci. B 8, 815–821.

Moradi, M.H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K.G., McEwan, J.C., 2012. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BMC Genet. 13, 10.

Nielsen, R., 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39, 197–218.

Nilsson-Ehle, P., Garfinkel, A.S., Schotz, M.C., 1980. Lipolytic enzymes and plasma lipoprotein metabolism. Annu. Rev. Biochem. 49, 667–693.

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

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., 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, X., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419 (2909), 832–837.
Schmid, A., Sutto, Z., Nlend, M.-C., Horvath, G., Schmid, N., Buck, J., Levin, L.R., Conner, G.E., Fregien, N., Salathe, M., 2007. Soluble adenylyl cyclase is localized to cilia and contributes to ciliary beat frequency regulation via production of cAMP. J. Gen. Physiol. 130, 99–109.
Sundström, H., Webster, M.T., Hans, E., 2004. Reduced variation on the chicken Z chromosome and the role of selective sweeps in sex chromosome evolution. Genetics 167 (1), 377–385.
Storchova, R., Divina, P., 2006. Nonrandom representation of sex-biased genes on chicken Z chromosome. J. Mol. Evol. 63 (5), 676–681.
Wang, H., 2013. Genome-wide association mapping including phenotypes from relatives without genotypes (Doctorate thesis). University of Georgia libraries.
Wu, X., Viveiros, M.M., Eppig, J.J., Bai, Y., Fitzpatrick, S.L., Matzuk, M.M., 2002. Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. Nat. Genet. 33, 187–191.