



# Pooled Sequencing Analysis of Geese (*Anser cygnoides*) Reveals Genomic Variations Associated With Feather Color

Shuang Ren<sup>1,2†</sup>, Guangqi Lyu<sup>1†</sup>, David M. Irwin<sup>3</sup>, Xin Liu<sup>1</sup>, Chunyu Feng<sup>1</sup>, Runhong Luo<sup>1</sup>, Junpeng Zhang<sup>1</sup>, Yongfeng Sun<sup>4</sup>, Songyang Shang<sup>1</sup>, Shuyi Zhang<sup>1</sup> and Zhe Wang<sup>1\*</sup>

<sup>1</sup> College of Animal Science and Veterinary Medicine, Shenyang Agricultural University, Shenyang, China, <sup>2</sup> College of Food Science, Shenyang Agricultural University, Shenyang, China, <sup>3</sup> Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada, <sup>4</sup> College of Animal Science and Technology, Jilin Agricultural University, Changchun, China

## OPEN ACCESS

### Edited by:

Luca Fontanesi,  
University of Bologna, Italy

### Reviewed by:

Andrea Talenti,  
University of Edinburgh,  
United Kingdom  
Ricardo Ventura,  
University of São Paulo, Brazil

### \*Correspondence:

Zhe Wang  
zwang@syau.edu.cn

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 06 January 2021

**Accepted:** 28 May 2021

**Published:** 18 June 2021

### Citation:

Ren S, Lyu G, Irwin DM, Liu X,  
Feng C, Luo R, Zhang J, Sun Y,  
Shang S, Zhang S and Wang Z (2021)  
Pooled Sequencing Analysis of Geese  
(*Anser cygnoides*) Reveals Genomic  
Variations Associated With Feather  
Color. *Front. Genet.* 12:650013.  
doi: 10.3389/fgene.2021.650013

During the domestication of the goose a change in its feather color took place, however, the molecular mechanisms responsible for this change are not completely understood. Here, we performed whole-genome resequencing on three pooled samples of geese (feral and domestic geese), with two distinct feather colors, to identify genes that might regulate feather color. We identified around 8 million SNPs within each of the three pools and validated allele frequencies for a subset of these SNPs using PCR and Sanger sequencing. Several genomic regions with signatures of differential selection were found when we compared the gray and white feather color populations using the  $F_{ST}$  and  $H_p$  approaches. When we combined previous functional studies with our genomic analyses we identified 26 genes (*KITLG*, *MITF*, *TYRO3*, *KIT*, *AP3B1*, *SMARCA2*, *ROR2*, *CSNK1G3*, *CCDC112*, *VAMP7*, *SLC16A2*, *LOC106047519*, *RLIM*, *KIAA2022*, *ST8SIA4*, *LOC106044163*, *TRPM6*, *TICAM2*, *LOC106038556*, *LOC106038575*, *LOC106038574*, *LOC106038594*, *LOC106038573*, *LOC106038604*, *LOC106047489*, and *LOC106047492*) that potentially regulate feather color in geese. These results substantially expand the catalog of potential feather color regulators in geese and provide a basis for further studies on domestication and avian feather coloration.

**Keywords:** goose, feather color, genome, pool-seq, SNP

## INTRODUCTION

Through more than 5000 years of constant artificial selection, domesticated geese have acquired a number of modifications to their appearance compared to their wild ancestors and relatives (Zeuner, 1963; Albarella, 2005). Most Asian and some European domestic goose breeds were derived from the swan goose (*Anser cygnoides*) (Buckland and Gérard, 2002). Domestication involved a complex set of metabolic, physiological and behavioral changes, including traits involving the liver, meat, eggs and feathers, but the most visible difference between wild and domestic swan geese is their feather coloration (Sossinka, 1982). Wild swan geese are characterized

by their iconic feathers with gray stripes (Figure 1A), while domestic swan geese have an all-white color appearance (Figures 1B,C; Gao et al., 2016).

Since feather coloration and patterns are prominent features in birds, and play essential roles in their survival, mechanisms that regulate the differentiation of feather color has been intensively studied (Abolins-Abols et al., 2018; Gao et al., 2018). Feather color is the consequence of two different, but related, physical processes, pigmentation and structural coloration, where pigmentation is the primary basis for the color diversity in animals (D'Alba et al., 2012). Melanins and carotenoids are widely distributed pigments in avian feathers and are the main contributors to the diversity of feather color in birds. The melanin content is usually higher than that of carotenoids, where studies have shown that the melanin content of feathers in swallows is four orders of magnitude greater than that of carotenoids (McGraw et al., 2004; McGraw, 2006; Delhey, 2015). The genetic control of melanogenesis in birds is achieved through genes that encode specific enzymes involved in melanin synthesis as well as other regulatory and structural proteins required for the distribution of melanin (Galván and Solano, 2016).

Investigations in some avian species have identified a limited number of genes involved in the mechanisms controlling feather coloration, however, only a few studies have focused on changes in feather color in geese or swans (Theron et al., 2001; Mundy et al., 2004; Bed'hom et al., 2012; Emaresi et al., 2013; Wang et al., 2014; Poelstra et al., 2015; Zhou et al., 2018; Wen et al., 2021). Melanic plumage polymorphisms in the lesser snow geese (*Anser caerulescens caerulescens*) and arctic skuas (*Stercorarius parasiticus*) correlate with changes in the copy number of variant *MC1R* alleles (Mundy et al., 2004). In the black and black-necked swans (*Cygnus atratus* and *C. melanocoryphus*), independently derived nucleotide substitutions in *MC1R*, which cause amino acid changes at important functional sites, have been identified that are consistent with increased *MC1R* activity and melanism pigment synthesis (Pointer and Mundy, 2008). In the domestic swan geese, three SNPs in *TYR* and one in *MITF* have been reported to be associated with white plumage (Wang et al., 2014). Recently, Wen et al. (2021) reported, in a genomic level examination of plumage color in domestic geese, an 18 bp deletion in an intron region of *KIT* (NW\_013185664.1, 11,785,718–11,785,736 bp) that was associated with white feather color.

Although *TYR*, *MITF* and *KIT* have been found to be associated with differences in feather coloration in domestic geese (Wang et al., 2014; Wen et al., 2021), a full understanding of the genetic basis of feather color formation in this species remains incomplete. With the unprecedented development of high-throughput sequencing, it has become possible to examine the genetic basis of differences in feather color at the genomic level (Zhou et al., 2018).

In this study, we performed whole-genome pooled sequencing (Pool-Seq) on three populations of swan geese with wild type and white-colored feathers. By identifying genomic regions that experienced selective sweeps, we aimed to identify genes that have experienced artificial selection and thus might explain the change in feather color in domesticated geese.

## MATERIALS AND METHODS

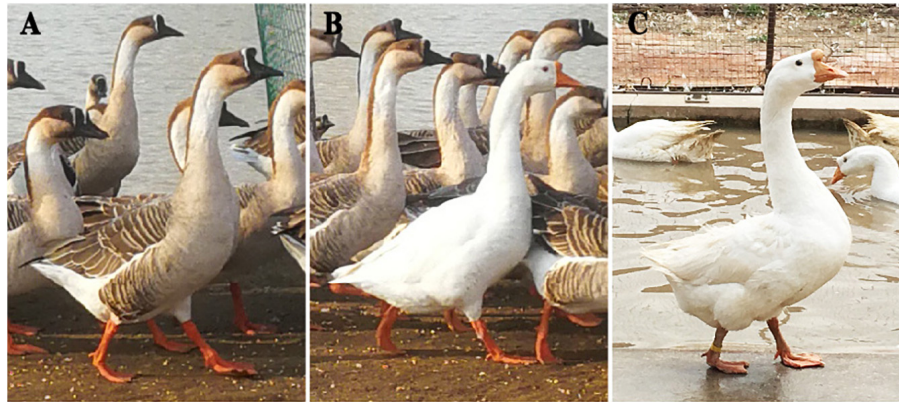
### Whole-Genome Pooled Sequencing of Goose DNA Samples

A total of 117 feral gray (*Anser cygnoides*, 60 females, 57 males, group gray; Figure 1A), 25 feral white (*Anser cygnoides*, 10 females, 15 males, group White\_1; Figure 1B), and 87 domesticated white (*Anser cygnoides domesticus*, 52 females, 35 males, group White\_2; Figure 1C) geese were sampled. These samples were collected from large populations to minimize genetic relationships. We choose our sample sizes for this Pool-seq study based on previous reports on Darwin's finch (sample size range 8–35) and monarch butterflies (sample size range 9–101) (Zhan et al., 2014; Lamichhaney et al., 2015). The accuracy of Pool-Seq increases with larger numbers of individuals included in the pool (Futschik and Schlötterer, 2010; Gautier et al., 2013). This suggests that our samples should be sufficient to identify SNPs and genes associated with feather color. A subset of SNPs identified from Pool-Seq were validated by Sanger sequencing to assess the accuracy of estimating allele frequencies (AFs) using Pool-Seq. Blood samples were collected by venipuncture. Gray and White\_1 geese were acquired from a population maintained at the Xianghai breeding base in Jilin city, Jilin province, China. White\_2 geese, belonging to the Huoyan breed, were obtained from the Liaoyang Animal Science Research Institute, Liaoning province, China. Geese in the group gray, with wild-type feather color, were the offspring of a mating between a population of male wild geese and female domestic geese. After several generations of breeding, a sub-population of feral white (White\_1) geese appeared among the feral gray geese. Although fed by humans, unlike the domesticated white geese (White\_2), both the feral gray and feral white geese possess flight abilities similar to those of wild geese, which is considered to be a signature of feralization (Gering et al., 2019).

Genomic DNA was extracted individually from blood samples of each goose using a Blood Genome DNA Extraction Kit (TIANGEN, DP348) following the manufacturer's instructions. Equimolar quantities (3 µg/ml) of DNA from each individual were pooled to establish the three sequencing libraries. The first pooled sample was from 117 feral gray geese, the second from 25 feral white geese and the third from 87 domestic white geese. The concentrations and purity of genomic DNA were checked before library construction. Libraries were generated via adapter ligation and DNA cluster preparation and subjected to 150 bp paired-end sequencing on an Illumina HiSeq 4000 platform. Sequencing depth of each library was at least 30×. Library construction and genome sequencing was conducted by the Beijing Genomics Institute Co., Ltd. (Shenzhen, China).

### Data Processing, Mapping and SNP Calling

We applied the PoolParty pipeline (Micheletti and Narum, 2018), which was designed for pool sequencing, to analyze the sequence data. The module PPAlign was used to align each read to the reference genome and for SNP calling. The parameters of module



**FIGURE 1** | Images of swan geese (*Anser cygnoides*) sequenced in this study showing their feather color phenotypes. **(A)** Feral gray geese (group gray) have the same feather phenotype as wild geese. **(B)** A feral white goose (group White\_1) living in a population of feral gray geese. **(C)** Domesticated white geese (group White\_2).

PPAlign were: “THREADZ = 32 BQUAL = 20 MAPQ = 5 SNPQ = 20 MINLENGTH = 25 INWIN = 3 MAF = 0.05 KMEM = Xmx4g MINDP = 10”. Briefly, BBDuk<sup>1</sup> was used to obtain clean data by trimming primer dimers and adapter sequences from the reads, discarding bases with quality lower than Q20 and reads with lengths less than 25 bp. BWA-MEM (Li and Durbin, 2009) was then used to map the clean data to the goose reference genome (*AnsCyg\_PRJNA183603\_v1.0*)<sup>2</sup> (Lu et al., 2015).

Prior to SNP calling, SAMBLASTER<sup>3</sup> was used to mark duplicate read pairs and compress the alignment to eliminate any bias generated during the PCR amplification for library preparation and/or sequencing (Faust and Hall, 2014). Aligned results were then sorted by Picard Tools<sup>4</sup> () and ambiguously mapped or unaligned reads were removed with SAMtools (Li et al., 2009). BCFtools (Li, 2011) was then used to call and filter the SNPs into a VCF file. Filtered alignments were combined in mpileup format for downstream analyses. SNPs with sequencing depth < 10 folds, quality < 20, minor allele frequency (MAF) < 0.05 or within 15 bp of indel were discarded.

## Variant Discovery and SNP Annotation

SNP annotation and the functional consequences of sequence variants were predicted using the Ensembl Variant Effect Predictor (VEP) tool using Ensembl database version 103 with the input VCF file (McLaren et al., 2016). Annotated results of VEP included transcripts, proteins, regulatory regions, and phenotype (McLaren et al., 2016). We grouped loss-of-function (LoF) variants into four categories (1, stop-gain and stop-loss; 2, frameshift indel; 3, donor and acceptor splice-site; and 4, initiator codon variants) (Sveinbjornsson et al., 2016). Marker coverage for

each gene included 10 kb of upstream and downstream flanking region (Potter et al., 2010). We focused on LoF variant annotation results for the downstream analysis.

## Sanger Sequencing Validation of SNP Allele Frequencies (AFs)

SNP AFs were calculated from the read depths of each allele in the Pool-Seq data. To confirm the accuracy of AFs estimated from the Pool-Seq data, we performed a Kendall W’s coordination coefficient test on a subset of the SNPs (28 loci) (Dodge and Commenges, 2006). Of these SNPs, 15 SNPs (SNP01, SNP06 and SNP11-23 in **Supplementary Table 1**) were selected as they had the lowest *P*-values in the comparison of the Gray and White\_2 groups by Fisher’s exact test based on read depth of alleles. Eight SNPs (SNP02-05 and SNP7-10 in **Supplementary Table 1**) were selected as they were adjacent to SNP01 and SNP06 and could be amplified with the same primer pairs used for them. Five SNPs (SNP24-28 in **Supplementary Table 1**) located in four genes (*KITLG*, *MITF*, *TYRO3*, and *KIT*) were also selected as these genes had previously been reported to be associated with the regulation of feather or coat color (Wehrle-Haller, 2003; Zhu et al., 2009; Zhou et al., 2018; Wu et al., 2019). The SNP alleles selected for validation were genotyped in all 229 individual geese by Sanger sequencing and the AFs were calculated from the genotype data. The two estimates of AFs, which were obtained from Pool-Seq and Sanger sequencing data, were compared using the Kendall W’s coordination coefficient test. Chi-square tests were performed to test the significance of the associations between the five SNPs in the color-related genes (*KITLG*, *MITF*, *TYRO3*, and *KIT*) and feather color phenotype. Primers used for the amplification of the selected SNPs are listed in **Supplementary Table 1**.

## Detection of Selective Sweeps

To accurately detect genomic regions in geese that had experienced selection during domestication and to estimate

<sup>1</sup><http://sourceforge.net/projects/bbmap/>

<sup>2</sup><https://www.ncbi.nlm.nih.gov/genome/?term=Anser+cygnoides>

<sup>3</sup><https://github.com/GregoryFaust/samblaster>

<sup>4</sup><http://broadinstitute.github.io/picard/>



the patterns of genetic diversity across the goose genome, we conducted selective sweep analyses including the fixed index ( $F_{ST}$ ) and pooled heterozygosity ( $H_p$ ) approaches (Rubin et al., 2010; Micheletti and Narum, 2018).  $F_{ST}$  in 10-kb non-overlapping sliding windows were calculated using the “fst-sliding.pl” module in Popoolation2 (Kofler et al., 2011), according to Weir and Cockerham’s method (Weir and Cockerham, 1984). The global parameters of  $F_{ST}$  approach were: “MINCOV = 10 MAXCOV = 100 MAF = 0.05.”  $H_p$  and negative Z-transformed  $H_p$  ( $-ZH_p$ ) were calculated using a custom python3 script in 10-kb non-overlapping sliding windows. The  $H_p$  approach determines, for each pool and SNP, the numbers of reads corresponding to the most ( $n_{MAJ}$ ) and least ( $n_{MIN}$ ) abundant alleles. For each window in each breed pool, the heterozygosity score of the pool was calculated as:

$$H_p = \frac{2 \sum n_{MAJ} \sum n_{MIN}}{(\sum n_{MAJ} + \sum n_{MIN})^2}$$

Where  $n_{MAJ}$  and  $n_{MIN}$  represent the numbers of reads corresponding to the most and least abundant allele. Individual  $H_p$  values were then Z-transformed as follows:

$$-ZH_p = -\frac{H_p - \mu H_p}{\sigma H_p}$$

Windows with less than 10 SNPs were discarded to avoid spurious signals. Windows located in the top 3% of the  $F_{ST}$  distribution and top 3% of the  $-ZH_p$  distribution were regarded as candidate regions for selective sweeps (Wang et al., 2015). Genes overlapping these regions were identified using Ensembl genome annotation.

Genomic regions that might have experienced selective sweeps were identified through three steps: (1) windows in the top 3% of the  $F_{ST}$  distributions of both the Gray vs. White\_1 and the Gray vs. White\_2 comparisons were identified; (2) windows in the top 3% of the  $-ZH_p$  distributions of both the White\_1 and the White\_2 populations were identified; (3) the intersection of region identified in (1) and (2) were considered to have experienced a selective sweep. Genes located in these overlapped regions might be involved in the change of goose feather color.

## Gene Ontology (GO) and KEGG Pathway Enrichment Analysis

To determine the possible function of genes that were located in the selective sweep regions, we identified orthologous human genes using the BioMart online tool<sup>5</sup>. The orthologous genes were then uploaded into the DAVID online tool to test for enrichment in gene ontology (GO) terms (Huang et al., 2009). KEGG pathway analysis was conducted using the online KOBAS tool (Xie et al., 2011). A Fisher’s exact test was then used to determine the significance of the enrichments of the GO terms and KEGG pathways, with a significant level of  $P < 0.05$ .

<sup>5</sup><http://www.ENSEMBL.org/biomart/martview/>

## RESULTS

### Statistics of the Genome Resequencing Data

A total of 148.26 Gb clean data was obtained from the three Pool-Seq libraries (Table 1). Mapping rates for the libraries varied between 98.14 and 98.23%, with the final effective mapping depths ranging from 44.09- to 44.13-fold. The Q20 rates for the three libraries were all over 98%. An average of 8,476,172 SNPs was identified in each library.

### Sanger Sequencing Validation

To assess the reliability of estimating allele frequencies of SNPs using the population genomic sequencing (Pool-Seq) data, we genotyped 28 SNPs from an average of 210 individuals using Sanger sequencing (Supplementary Tables 2, 3). AFs calculated from the Sanger sequencing data, based on individual amplifications and sequencing, were in accord with the AFs calculated from the Pool-Seq data. Kendall W’s coordination coefficients for the comparisons of the AFs estimated from the Pool-Seq and Sanger genotypes for the Gray, White\_1 and White\_2 populations were 0.96, 0.97 and 0.94 ( $P < 0.05$ ), respectively, showing that there is a good concordance between the results obtained using the two different methods.

Among the 28 SNPs examined above, five were SNPs that are located in four genes (*KITLG*, *MITF*, *TYRO3*, and *KIT*) previously reported to be associated with feather or coat color (Wehrle-Haller, 2003; Zhu et al., 2009; Zhou et al., 2018; Wu et al., 2019). Of these five SNPs, two are located in the 3’ UTR of *KITLG* (NW\_013185706.1: G232853A and NW\_013185706.1: C232854T), one in the 5’ UTR of *MITF* (NW\_013185692.1: G4400553C), one in the *TYRO3* (S772G) coding region and one in the *KIT* (T887A) coding region (Table 2). Results from a Chi-square test showed an extremely significant association between the SNP genotypes and feather color phenotypes ( $P < 0.001$ ) for these SNPs.

### Selective Sweep Analysis

$H_p$  and  $-ZH_p$  distributions are presented in Supplementary Figure 1. The selective sweep analyses identified (1) 317 regions in the top 3% of the  $F_{ST}$  distributions of the intersection of the Gray vs. White\_1 ( $F_{ST}$  value, mean = 0.119, range 0.095–0.457) and the Gray vs. White\_2 ( $F_{ST}$  value, mean = 0.258, range 0.202–0.727) comparisons and (2) 253 regions in the top 3% of the  $-ZH_p$  distributions of the intersection of the

**TABLE 1** | Summary statistics of the clean data from whole-genome resequencing.

Parameter	Gray	White_1	White_2
Clean data (Gb)	49.42	49.40	49.44
Reads (M)	329.45	329.30	329.59
Map reads rate (%)	98.14	98.20	98.23
Q20 rate (%)	98.08	98.14	98.17
Sequencing depth	44.11	44.09	44.13
Total SNPs	8,785,296	8,680,731	7,962,489

**TABLE 2** | Genotypes of the five PCR verified loci.

SNP information <sup>a</sup>	Gene	Genotype	Numbers of individuals <sup>b</sup>			$\chi^2$ value
			Gray	White_1	White_2	
NW_013185706.1: 3'UTR_G232853A	KITLG	GG	37	22 (88%)	84 (100%)	102.89 ( $P < 0.001$ )
		GA	77 (68%)	3	0	
		AA	0	0	0	
NW_013185706.1: 3'UTR_C232854T	KITLG	CC	37	22 (88%)	84 (100%)	102.89 ( $P < 0.001$ )
		CT	77 (68%)	3	0	
		TT	0	0	0	
NW_013185692.1: 5'UTR_G4400553C	MITF	GG	83 (74%)	6	14	80.44 ( $P < 0.001$ )
		GC	29	17 (68%)	49 (59%)	
		CC	0	2	20	
NW_013185657.1: cds_A2638G:S772G	TYRO3	AA	31	20 (80%)	86 (100%)	106.57 ( $P < 0.001$ )
		AG	76 (71%)	5	0	
		GG	0	0	0	
NW_013185664.1: cds_A2659G:T887A	KIT	AA	0	0	0	93.80 ( $P < 0.001$ )
		AG	67 (66%)	2	0	
		GG	35	23 (92%)	78 (100%)	

<sup>a</sup>SNP information is displayed in the following format: NCBI GenBank accession number: nucleotide position in the goose reference genome (cds, 5'UTR or 3'UTR) + nucleotide in the reference genome (reference nucleotide) + SNP position in the transcript + mutant nucleotide: amino acid in the goose reference transcript + codon position in the transcript + amino acid mutation or no change.

<sup>b</sup>Numbers of individuals (Gray| White\_1| White\_2) for which reliable genotypes could be inferred from the Sanger sequencing data. The percentage in parenthesis is the percentage of the genotype in the number of verified individuals.

White\_1 (−ZHp value, mean = 2.835, range 2.014–4.952) and White\_2 (−ZHp value, mean = 2.904, range 2.317–3.755) populations (Figure 2). A total of 99 genes were identified in the 317 regions identified by the  $F_{ST}$  distributions and 103 genes identified in the 253 regions identified by the −ZHp distributions (Supplementary Tables 4, 5). Among the 99 genes identified from the  $F_{ST}$  distributions, four (*SLC16A2*, *AP3B1*, *SMARCA2*, and *VAMP7*) have previously been associated with animal coloration (Table 3). Similarly, 5 of the 103 genes from the −ZHp distributions (*SLC16A2*, *ROR2*, *CSNK1G3*, *CCDC112*, and *VAMP7*) were previously associated with animal coloration (Table 3).

More importantly, 27 regions, which included 17 genes, were found in the overlap of the 317  $F_{ST}$  and 253 −ZHp top 3% regions, including two, *SLC16A2* and *VAMP7* that had previously been associated with animal coloration (Table 4). The 15 novel genes identified here are: *LOC106038556*, *LOC106038575*, *LOC106038574*, *LOC106038594*, *LOC106038573*, *RLIM*, *LOC106038604*, *KIAA2022*, *ST8SIA4*, *LOC106044163*, *TRPM6*, *TICAM2*, *LOC106047489*, *LOC106047492*, and *LOC106047519* (Table 4). LoF (loss of function) variants were found for 9 of the genes (8 detected by both the  $F_{ST}$  and the *Hp* approaches (*LOC106038574*, *LOC106038604*, *LOC106044163*, *TICAM2*, *LOC106047489*, *VAMP7*, *LOC106047492*, and *LOC106047519*) and one by only *Hp* (*CCDC112*) in the selective sweep regions of geese with different feather colors (or in the 10 kb region upstream and downstream of the sweep regions) (Supplementary Table 6).

## GO and KEGG Pathway Enrichment Analysis

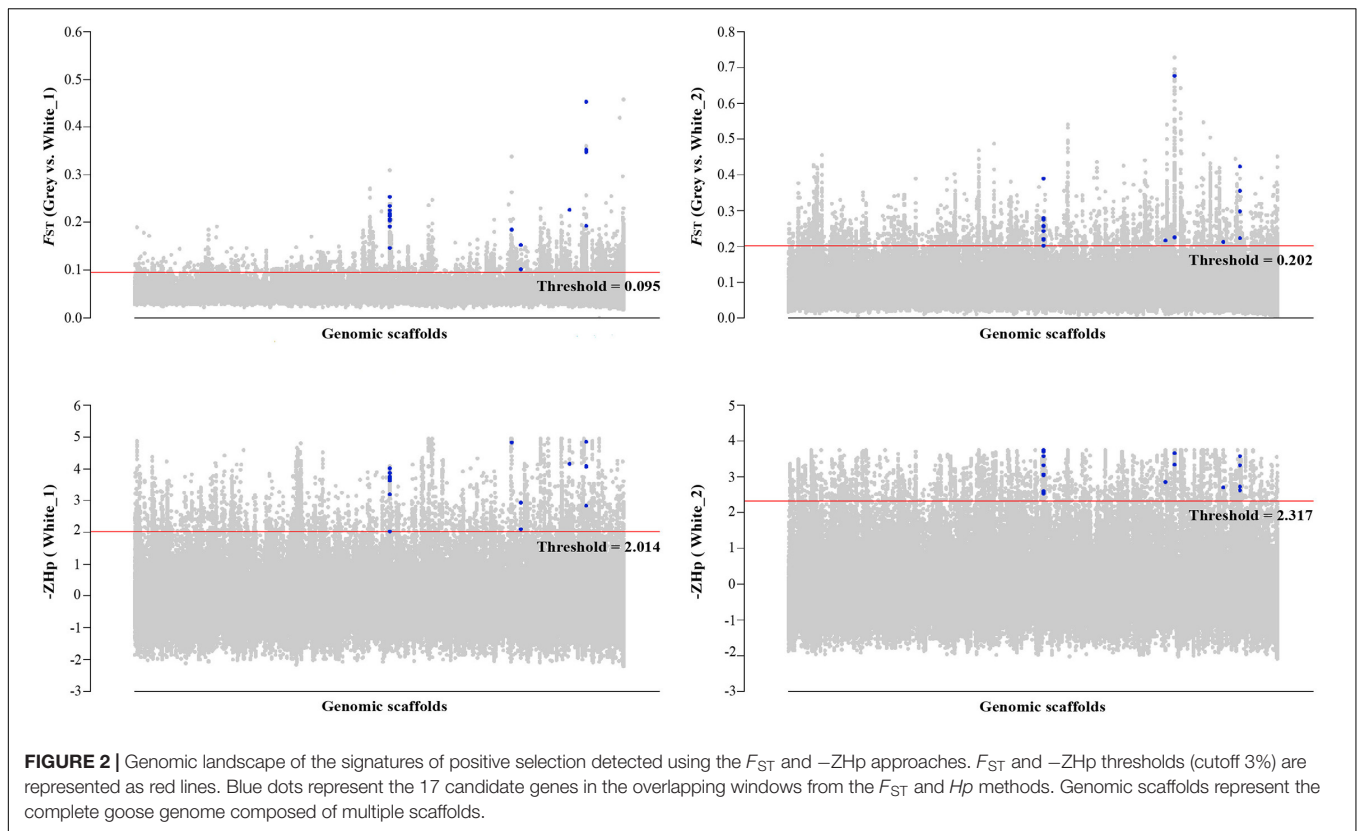
We conducted GO and KEGG pathway analyses of the 17 genes identified in the selective sweep regions. These genes were found

to be significantly enriched for the GO term “late endosomal membrane category” (GO: 0031902,  $P < 0.05$ ) and in three pathways, “RNA polymerase,” “SNARE interactions in vesicular transport” and “cytosolic DNA-sensing pathway” ( $P < 0.05$ , Supplementary Table 7).

## DISCUSSION

In this study, we performed whole-genome Pool-Seq on three populations of geese with two different colors of feathers to identify SNPs, and genes that might be responsible for these differing phenotypes. The Kendall  $W$ 's coefficients for the AFs calculated from the Pool-Seq and Sanger sequencing data indicated a good correlation between them, which suggests that our Pool-Seq data is adequate for identifying loci that are differentiated between the goose phenotypes. Selective sweep analyses of this SNP data was used to identify genomic regions that show signatures of selection during the domestication of geese. This led to the identification of 17 genes located in candidate regions identified by both the  $F_{ST}$  and *Hp* approaches, suggesting a high probability that selection occurred on these genes and that they might be associated with the change in feather color seen in these geese. VEP annotation of these 17 genes identified eight with loss-of-function (LoF) alleles potentially involved in regulating feather color.

Among the 17 identified genes (Table 4), three (*VAMP7*, *SLC16A2*, and *LOC106047519*) have previously been associated with the regulation of coat color in animals (Imokawa et al., 1992; Yatsu et al., 2013; Baxter et al., 2019). *VAMP7*, vesicle associated membrane protein 7, is localized to Tyrp1-containing vesicles/organelles and acts as part of the SNARE machinery with syntaxin-3 and SNAP23 on melanosomes to regulate Tyrp1 transport in mouse melanocytes (Yatsu et al., 2013).



*VAMP7* may play a key role in melanin formation and thus influence goose feather color. *SLC16A2* has an effect on pigmentation phenotypes in the zebrafish, and has the GO

**TABLE 3 |** Genes associated with animal coloration that overlap with the selected regions under detected by the  $F_{ST}$  or  $H_p$  approach.

Method	Scaffold <sup>a</sup>	Gene symbol	Summary of gene function
$F_{ST}$ (Top 3%)	NW_013185770.1	<i>AP3B1</i>	Melanin formation (Jing et al., 2014)
	NW_013185722.1	<i>SLC16A2</i>	Pigment-related (Baxter et al., 2019)
	NW_013185909.1	<i>SMARCA2</i>	Melanin formation (Mehrotra et al., 2014)
	NW_013185915.1	<b><i>VAMP7</i></b>	Melanin formation (Yatsu et al., 2013)
$H_p$ (Top 3%)	NW_013185722.1	<i>SLC16A2</i>	Pigment-related (Baxter et al., 2019)
	NW_013185840.1	<i>ROR2</i>	Melanin formation (O'Connell et al., 2013)
	NW_013185881.1	<i>CSNK1G3</i>	Pigment-related (Al Robaee et al., 2020)
	NW_013185883.1	<b><i>CCDC112</i></b>	Pigment-related (Tian et al., 2014)
	NW_013185915.1	<b><i>VAMP7</i></b>	Melanin formation (Yatsu et al., 2013)

<sup>a</sup>AnsCyg\_PRJNA183603\_v1.0 primary assembly.

Gene names in bold have LoF variation according to VEP annotation.

term “pigmentation” annotated in the Zebrafish Information Network<sup>6</sup> database (Baxter et al., 2019). *LOC106047519* belongs to the ETB-R gene family, which also includes the Endothelin B receptor (*EDNRB*), and has been described as an *EDNRB*-like gene (Kanehisa, 1997; Kanehisa and Goto, 2000). *EDNRB* is reported to be associated with the development of cells of the melanocytic lineage (Imokawa et al., 1992), suggesting that *LOC106047519* might also perform a function similar to *EDNRB* to regulate feather color. Our VEP analysis identified LoF mutations in *VAMP7* and *LOC106047519*, but not in *SLC16A2*. These results suggest that *VAMP7* and *LOC106047519* might not only regulate pigmentation in the previously investigated animals but also play a role in the change in feather color in the goose.

Using either the  $F_{ST}$  or the  $H_p$  approach we identified five other genes (*AP3B1*, *SMARCA2*, *ROR2*, *CSNK1G3*, and *CCDC112*) that have been reported to be associated with animal coloration (Table 3). Substitutions in *AP3B1* cause distinct phenotypes in the pigmented cells in mouse eyes and possibly plays a role in organelle biogenesis associated with melanosomes (Jing et al., 2014). *SMARCA2*, a member of the SWI/SNF family, is involved in melanocyte differentiation and melanoma (Mehrotra et al., 2014; Markiewicz and Idowu, 2020). *ROR2* is involved in the formation of melanoma in humans, suggesting a role in melanin formation (O'Connell et al., 2013). Expression of *CSNK1G3*, a gene related to human

<sup>6</sup><https://zfin.org>

**TABLE 4** | 17 genes found in the overlapping regions identified by both the  $F_{ST}$  and the  $H_p$  differentiation approaches.

Scaffold <sup>a</sup>	Position <sup>b</sup>	Gene symbol	Gene description
NW_013185722.1	261,870	<i>LOC106038556</i>	Homeobox protein CDX-4-like
NW_013185722.1	288,464	<i>LOC106038575</i>	Uncharacterized LOC106038575
NW_013185722.1	301,490	<b><i>LOC106038574</i></b>	Pre-mRNA 3'-end-processing factor FIP1-like
NW_013185722.1	338,779	<i>LOC106038594</i>	Ligand of Numb protein X 2-like
NW_013185722.1	371,156	<i>LOC106038573</i>	Uncharacterized LOC106038573
NW_013185722.1	502,677	<i>SLC16A2</i>	Solute carrier family 16 member 2
NW_013185722.1	517,620	<i>RLIM</i>	Ring finger protein, LIM domain interacting
NW_013185722.1	537,179	<b><i>LOC106038604</i></b>	Uncharacterized LOC106038604
NW_013185722.1	548,033	<i>KIAA2022</i>	Neurite extension and migration factor
NW_013185807.1	123,1519	<i>ST8SIA4</i>	ST8 alpha-N-acetylneuraminide alpha-2,8-sialyltransferase 4
NW_013185817.1	975,069	<b><i>LOC106044163</i></b>	Protein convertase subtilisin/kexin type 5-like
NW_013185817.1	143,0086	<i>TRPM6</i>	Transient receptor potential cation channel subfamily M member 6
NW_013185883.1	304,453	<b><i>TICAM2</i></b>	Toll like receptor adaptor molecule 2
NW_013185915.1	689,355	<b><i>LOC106047489</i></b>	SLAIN motif-containing protein-like
NW_013185915.1	717,937	<b><i>VAMP7</i></b>	Vesicle associated membrane protein 7
NW_013185915.1	737,829	<b><i>LOC106047492</i></b>	DNA-directed RNA polymerases I and III subunit RPAC2-like
NW_013185915.1	745,589	<b><i>LOC106047519</i></b>	Endothelin B receptor-like

<sup>a</sup>Same as **Table 3**.

<sup>b</sup>Start position of the gene in the genome.

Eight gene names in bold have LoF variants according to VEP annotation.

vitiligo, is significantly reduced in C57BL/6 black mice with tyrosinase-induced depigmented skin (Ocampo-Candiani et al., 2018; Al Robaee et al., 2020). *CCDC112* regulates pigmentation and the expression level of this gene differs between Silkie and White Leghorn chickens (Tian et al., 2014). We found a LoF mutation in *CCDC112* in gray that might partly explain the difference in feather color in geese. Our results suggest that it is possible that these five genes also affect feather color in geese.

We also focused on four genes (*KITLG*, *MITF*, *TYRO3*, and *KIT*) that were previously reported to be associated with feather color (Wehrle-Haller, 2003; Zhu et al., 2009; Zhou et al., 2018; Wu et al., 2019). The SNP genotypes for these genes were also validated by Sanger sequencing (**Table 2**). A change in an untranslated region (UTRs) can lead to changes in the

expression of genes (Barrett et al., 2012). Here, we identified three SNPs, two located in the 3' UTR of *KITLG* and one in the 5' UTR of *MITF*, which are significantly associated with feather color phenotypes in our geese. This suggests that these three SNPs affect the expression of *KITLG* and *MITF* resulting in a change in feather color. Non-synonymous mutations are more likely to affect the biological function of a gene. Here, we identified two non-synonymous substitutions in *KIT* (T887A) and *TYRO3* (S772G) that are significantly associated with feather color phenotypes, indicating that they may regulate goose feather color.

GO and KEGG enrichment analyses of the 17 candidate genes in the most significant sweep areas are significantly enriched in the GO term late endosome membrane (**Supplementary Table 7**). Two of the candidate genes (*VAMP7* and *TICAM2*) are associated with this term. We also identified three pathways (RNA polymerase, SNARE interactions in vesicular transport and cytosolic DNA-sensing pathway) that are significantly enriched, where *VAMP7* is also involved with SNARE interactions in vesicular transport. Although the enriched GO term and the pathways do not seem to directly correlate with animal coloration, it is still possible that the genes involved in them could regulate feather coloration in geese.

In conclusion, we identified 26 genes (17 detected by both the  $F_{ST}$  and  $H_p$  approaches, five by either  $F_{ST}$  or  $H_p$  and four previously reported color-related genes) from our genomic Pool-Seq data that might be responsible for the change in feather color that occurred during the domestication of geese (*Anser cygnoides*). Among these 26 genes, 12 have previously been found to be associated with animal coloration in other studies. The roles of the other genes in feather coloration requires further investigation. Additional studies, including functional experimentation, are needed to confirm the roles of these genes, and the consequence of the mutations caused by the SNPs, on phenotypic variation in feather color in geese.

## DATA AVAILABILITY STATEMENT

Whole-genome sequencing data reported in this study were deposited into the NCBI Sequence Read Archive under the accession number PRJNA532466 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA532466/>).

## ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care and Use Committee of Shenyang Agricultural University. Written informed consent was obtained from the owners for the participation of their animals in this study.

## AUTHOR CONTRIBUTIONS

SR, XL, CF, and RL performed the experiments. SR and GL analyzed the data. JZ, YS, and SZ collected the samples. SR, GL, SS, DI, and ZW wrote the manuscript. SZ and ZW designed the



study and supervised the work. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by grants from the National Natural Science Foundation of China (No. 31672274), the Organization Department of Liaoning Provincial Committee

(No. XLYC1907018), and the Educational Department of Liaoning Province of China (Climbing Scholar).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.650013/full#supplementary-material>

## REFERENCES

- Abolins-Abols, M., Kornobis, E., Ribeca, P., Wakamatsu, K., Peterson, M. P., Ketterson, E. D., et al. (2018). Differential gene regulation underlies variation in melanic plumage coloration in the dark-eyed Junco (*Junco hyemalis*). *Mol. Ecol.* 27, 4501–4515. doi: 10.1111/mec.14878
- Al Robaee, A. A., Alzolibani, A. A., and Rasheed, Z. (2020). Autoimmune response against tyrosinase induces depigmentation in C57BL/6 black mice. *Autoimmunity* 53, 459–466. doi: 10.1080/08916934.2020.1836489
- Albarella, U. (2005). Alternate fortunes? The role of domestic ducks and geese from Roman to Medieval times in Britain. *Doc. Archaeobiol.* 3, 249–258.
- Barrett, L. W., Fletcher, S., and Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* 69, 3613–3634.
- Baxter, L. L., Watkins-Chow, D. E., Pavan, W. J., and Loftus, S. K. (2019). A curated gene list for expanding the horizons of pigmentation biology. *Pigment Cell Melanoma Res.* 32, 348–358. doi: 10.1111/pcmr.12743
- Bed'hom, B., Vaez, M., Coville, J.-L., Gourichon, D., Chastel, O., Follet, S., et al. (2012). The lavender plumage colour in Japanese quail is associated with a complex mutation in the region of MLPH that is related to differences in growth, feed consumption and body temperature. *BMC Genomics* 13:442. doi: 10.1186/1471-2164-13-442
- Buckland, R. B., and Gérard, G. (2002). "Origins and Breeds of Domestic Geese," in *Goose Production*, ed. R. Buckland. (Rome: Food & Agriculture Org).
- D'Alba, L., Kieffer, L., and Shawkey, M. D. (2012). Relative contributions of pigments and biophotonic nanostructures to natural color production: a case study in budgerigar (*Melopsittacus undulatus*) feathers. *J. Exp. Biol.* 215, 1272–1277. doi: 10.1242/jeb.064907
- Delhey, K. (2015). The colour of an avifauna: a quantitative analysis of the colour of Australian birds. *Sci. Rep.* 5:18514. doi: 10.1038/srep18514
- Dodge, Y., and Commenges, D. (2006). *The Oxford Dictionary of Statistical Terms*. New York, NY: Oxford University Press on Demand.
- Emaresi, G., Ducrest, A. L., Bize, P., Richter, H., Simon, C., and Roulin, A. (2013). Pleiotropy in the melanocortin system: expression levels of this system are associated with melanogenesis and pigmentation in the tawny owl (*Strix aluco*). *Mol. Ecol.* 22, 4915–4930. doi: 10.1111/mec.12438
- Faust, G. G., and Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30, 2503–2505. doi: 10.1093/bioinformatics/btu314
- Futschik, A., and Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186, 207–218. doi: 10.1534/genetics.110.114397
- Galván, I., and Solano, F. (2016). Bird integumentary melanins: biosynthesis, forms, function and evolution. *Int. J. Mol. Sci.* 17:520. doi: 10.3390/ijms17040520
- Gao, G., Xu, M., Bai, C., Yang, Y., Li, G., Xu, J., et al. (2018). Comparative genomics and transcriptomics of *Chrysolophus* provide insights into the evolution of complex plumage coloration. *Gigascience* 7:giy113. doi: 10.1093/gigascience/gy113
- Gao, G., Zhao, X., Li, Q., He, C., Zhao, W., Liu, S., et al. (2016). Genome and metagenome analyses reveal adaptive evolution of the host and interaction with the gut microbiota in the goose. *Sci. Rep.* 6:32961. doi: 10.1038/srep32961
- Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., et al. (2013). Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol. Ecol.* 22, 3766–3779. doi: 10.1111/mec.12360
- Gering, E., Incorvaia, D., Henriksen, R., Conner, J., Getty, T., and Wright, D. (2019). Getting back to nature: feralization in animals and plants. *Trends Ecol. Evol.* 34, 1137–1151. doi: 10.1016/j.tree.2019.07.018
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (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
- Imokawa, G., Yada, Y., and Miyagishi, M. (1992). Endothelins secreted from human keratinocytes are intrinsic mitogens for human melanocytes. *J. Biol. Chem.* 267, 24675–24680.
- Jing, R., Dong, X., Li, K., Yan, J., Chen, X., and Feng, L. (2014). The Ap3b1 gene regulates the ocular melanosome biogenesis and tyrosinase distribution differently from the Hps1 gene. *Exp. Eye Res.* 128, 57–66. doi: 10.1016/j.exer.2014.08.010
- Kanehisa, M. (1997). A database for post-genome analysis. *Trends Genet.* 13, 375–376.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kofler, R., Pandey, R. V., and Schlötterer, C. (2011). PoolSeq: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27, 3435–3436. doi: 10.1093/bioinformatics/btr589
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518, 371–375. doi: 10.1038/nature14181
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lu, L., Chen, Y., Wang, Z., Li, X., Chen, W., Tao, Z., et al. (2015). The goose genome sequence leads to insights into the evolution of waterfowl and susceptibility to fatty liver. *Genome Biol.* 16:89. doi: 10.1186/s13059-015-0652-y
- Markiewicz, E., and Idowu, O. C. (2020). Melanogenic difference consideration in ethnic skin type: a balance approach between skin brightening applications and beneficial sun exposure. *Clin. Cosmet. Invest. Dermatol.* 13:215. doi: 10.2147/CCID.S245043
- McGraw, K. J. (2006). Pterins, porphyrins, and Psittacofulvins. *Bird Color. Mech. Meas.* 1, 354–398.
- McGraw, K. J., Safran, R. J., Evans, M. R., and Wakamatsu, K. (2004). European barn swallows use melanin pigments to color their feathers brown. *Behav. Ecol.* 15, 889–891. doi: 10.1093/beheco/arh109
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17, 1–14. doi: 10.1186/s13059-016-0974-4
- Mehrotra, A., Mehta, G., Aras, S., Trivedi, A., and de la Serna, I. (2014). SWI/SNF chromatin remodeling enzymes in melanocyte differentiation and melanoma. *Crit. Rev. Eukaryot. Gene Expr.* 24, 151–161. doi: 10.1615/CritRevEukaryotGeneExpr.v24.i2



- Micheletti, S. J., and Narum, S. R. (2018). Utility of pooled sequencing for association mapping in nonmodel organisms. *Mol. Ecol. Resour.* 18, 825–837. doi: 10.1111/1755-0998.12784
- Mundy, N. I., Badcock, N. S., Hart, T., Scribner, K., Janssen, K., and Nadeau, N. J. (2004). Conserved genetic basis of a quantitative plumage trait involved in mate choice. *Science* 303, 1870–1873. doi: 10.1126/science.1093834
- Ocampo-Candiani, J., Salinas-Santander, M., Trevino, V., Ortiz-López, R., Ocampo-Garza, J., and Sanchez-Dominguez, C. N. (2018). Evaluation of skin expression profiles of patients with vitiligo treated with narrow-band UVB therapy by targeted RNA-seq. *An. Bras. Dermatol.* 93, 843–851. doi: 10.1590/abd1806-4841.20187589
- O'Connell, M. P., Marchbank, K., Webster, M. R., Valiga, A. A., Kaur, A., Vultur, A., et al. (2013). Hypoxia induces phenotypic plasticity and therapy resistance in melanoma via the tyrosine kinase receptors ROR1 and ROR2. *Cancer Discov.* 3, 1378–1393. doi: 10.1158/2159-8290
- Poelstra, J. W., Vijay, N., Hoepfner, M. P., and Wolf, J. B. W. (2015). Transcriptomics of colour patterning and coloration shifts in crows. *Mol. Ecol.* 24, 4617–4628. doi: 10.1111/mec.13353
- Pointer, M. A., and Mundy, N. I. (2008). Testing whether macroevolution follows microevolution: Are colour differences among swans (*Cygnus*) attributable to variation at the MC1R locus? *BMC Evol. Biol.* 8:249. doi: 10.1186/1471-2148-8-249
- Potter, C., Cordell, H. J., Barton, A., Daly, A. K., Hyrich, K. L., Mann, D. A., et al. (2010). Association between anti-tumour necrosis factor treatment response and genetic variants within the TLR and NF $\kappa$ B signalling pathways. *Ann. Rheum. Dis.* 69, 1315–1320. doi: 10.1136/ard.2009.117309
- Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587–591. doi: 10.1038/nature08832
- Sossinka, R. (1982). Domestication in birds. *Avian Biol.* 6, 373–403.
- Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S. A., Oddsson, A., Måsson, G., et al. (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* 48:314. doi: 10.1038/ng.3507
- Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R. E., and Mundy, N. I. (2001). The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit. *Coereba Flaveola. Curr. Biol.* 11, 550–557. doi: 10.1016/S0960-9822(01)00158-0
- Tian, M., Hao, R., Fang, S., Wang, Y., Gu, X., Feng, C., et al. (2014). Genomic regions associated with the sex-linked inhibitor of dermal melanin in Silkie chicken. *Front. Agric. Sci. Eng.* 1:242. doi: 10.15302/J-FASE-2014018
- Wang, C., Wang, H., Zhang, Y., Tang, Z., Li, K., and Liu, B. (2015). Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. *Mol. Ecol. Resour.* 15, 414–424. doi: 10.1111/1755-0998.12311
- Wang, Y., Li, S. M., Huang, J., Chen, S. Y., and Liu, Y. P. (2014). Mutations of TYR and MITF genes are associated with plumage colour phenotypes in geese. *Asian Austr. J. Anim. Sci.* 27, 778–783. doi: 10.5713/ajas.2013.13350
- Wehrle-Haller, B. (2003). The role of Kit-ligand in melanocyte development and epidermal homeostasis. *Pigment Cell Res.* 16, 287–296. doi: 10.1034/j.1600-0749.2003.00055.x
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution (N.Y.)* 38, 1358–1370. doi: 10.2307/2408641
- Wen, J., Shao, P., Chen, Y., Wang, L., Lv, X., Yang, W., et al. (2021). Genomic scan revealed KIT gene underlying white/gray plumage color in Chinese domestic geese. *Anim. Genet.* 52, 356–360. doi: 10.1111/age.13050
- Wu, Z., Deng, Z., Huang, M., Hou, Y., Zhang, H., Chen, H., et al. (2019). Whole genome resequencing identifies KIT new alleles that affect coat color phenotypes in pigs. *Front. Genet.* 10:218. doi: 10.3389/fgene.2019.00218
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483
- Yatsu, A., Ohbayashi, N., Tamura, K., and Fukuda, M. (2013). Syntaxin-3 is required for melanosomal localization of Tyrp1 in melanocytes. *J. Invest. Dermatol.* 133, 2237–2246. doi: 10.1038/jid.2013.156
- Zeuner, F. E. (1963). *A History of Domesticated Animals*. London: Hutchinson.
- Zhan, S., Zhang, W., Niitepold, K., Hsu, J., Haeger, J. F., Zalucki, M. P., et al. (2014). The genetics of monarch butterfly migration and warning colouration. *Nature* 514, 317–321. doi: 10.1038/nature13812
- Zhou, Z., Li, M., Cheng, H., Fan, W., Yuan, Z., Gao, Q., et al. (2018). An intercross population study reveals genes associated with body size and plumage color in ducks. *Nat. Commun.* 9:2648. doi: 10.1038/s41467-018-04868-4
- Zhu, S., Wurdak, H., Wang, Y., Galkin, A., Tao, H., Li, J., et al. (2009). A genomic screen identifies TYRO3 as a MITF regulator in melanoma. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17025–17030. doi: 10.1073/pnas.0909292106

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ren, Lyu, Irwin, Liu, Feng, Luo, Zhang, Sun, Shang, Zhang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.