Artificial selection footprints in indigenous and commercial chicken genomes

Siwen Wu[2†], Tengfei Dou[1†], Kun Wang[1†], Sisi Yuan[2†], Shixiong Yan[1†], Zhiqiang Xu[1], Yong Liu[1], Zonghui Jian[1], Jingying Zhao[1], Rouhan Zhao[1], Hao Wu[1], Dahai Gu[1], Lixian Liu[1], Qihua Li[1], Dong-Dong Wu[3,4], Changrong Ge[1*], Zhengchang Su[2*] and Junjing Jia[1*]

Abstract

Background Although many studies have been done to reveal artificial selection signatures in commercial and indigenous chickens, a limited number of genes have been linked to specific traits. To identify more trait-related artificial selection signatures and genes, we re-sequenced a total of 85 individuals of five indigenous chicken breeds with distinct traits from Yunnan Province, China.

Results We found 30 million non-redundant single nucleotide variants and small indels (< 50 bp) in the indigenous chickens, of which 10 million were not seen in 60 broilers, 56 layers and 35 red jungle fowls (RJFs) that we compared with. The variants in each breed are enriched in non-coding regions, while those in coding regions are largely tolerant, suggesting that most variants might affect *cis*-regulatory sequences. Based on 27 million bi-allelic single nucleotide polymorphisms identified in the chickens, we found numerous selective sweeps and affected genes in each indigenous chicken breed and substantially larger numbers of selective sweeps and affected genes in the broilers and layers than previously reported using a rigorous statistical model. Consistent with the locations of the variants, the vast majority (~ 98.3%) of the identified selective sweeps overlap known quantitative trait loci (QTLs). Meanwhile, 74.2% known QTLs overlap our identified selective sweeps. We confirmed most of previously identified trait-related genes and identified many novel ones, some of which might be related to body size and high egg production traits. Using RT-qPCR, we validated differential expression of eight genes (*GHR, GHRHR, IGF2BP1, OVALX, ELF2, MGARP, NOCT, SLC25A15*) that might be related to body size and high egg production traits in relevant tissues of relevant breeds.

Conclusion We identify 30 million single nucleotide variants and small indels in the five indigenous chicken breeds, 10 million of which are novel. We predict substantially more selective sweeps and affected genes than previously reported in both indigenous and commercial breeds. These variants and affected genes are good candidates for further experimental investigations of genotype-phenotype relationships and practical applications in chicken breeding programs.

Keywords Chicken, Domestication, Evolution, Selective sweep, SNPs, Permutation

†Siwen Wu, Tengfei Dou and Kun Wang contributed equally to this work.

*Correspondence:
Changrong Ge
gcrzal@126.com
Zhengchang Su
zcsu@uncc.edu
Junjing Jia
junjingli2009@hotmail.com
Full list of author information is available at the end of the article

Wu *et al. BMC Genomics*   (2024) 25:428

Page 2 of 20

Background

Chicken (*Gallus gallus*) has been domesticated by human for about 8000 years [1], and multiple lines of evidence show that red jungle fowl (RJF) is the major ancestor of domestic chicken all over the world [1–3]. Artificial selection has resulted in numerous chicken breeds with distinct traits in different parts of the world for various purposes, including meat and egg production as well as recreation and ornament. Particularly, intensive systematic artificial selections carried out in a few companies in the last decays have led to highly production-efficient commercial broiler and layer lines used all over the world. Understanding the genetic basis of distinct traits of traditionally bred indigenous chicken as well as of commercial broilers and layers is crucial to guide breeding programs to further improve the chicken welfare [4]. Besides commercial lines, indigenous chicken breeds are also excellent model systems to study the relationships between genotypes and phenotypes [5]. Indeed, many studies have been done to reveal artificial selection signatures on commercial broilers and layers [6–8] as well as on indigenous chickens [9–17]. These studies have identified genes or quantitative trait loci (QTLs) related to specific traits such as body size [10, 11, 18–23], meat quality [24], egg production [25], feathering [26–33], plumage color [34–37], skin color [38], behavior [39], immunity [40], crest shapes [41], bone traits [42–44], rumpless trait [45–48], and polydactyly [49, 50]. Yunnan, a southwest province of China, is one of the major centers where domestic chickens arise [3], and numerous chicken breeds have been raised in mountainous areas there. Among these indigenous chicken breeds are Daweishan, Hu, Piao, Wuding and Nine-claw chicken, each with distinct traits. Specifically, Daweishan chickens have a miniature body size (0.5 ~ 0.8 kg for female and 0.8 ~ 1.2 kg for male adults); Hu chickens have a large body size (3 kg for female and 6 kg for male adults) with extraordinarily stout legs; Piao chickens have a short tail (a rumpless phenotype); Wuding chickens have a relatively large body size with colorful feathers and thick fat; and Nine-claw chickens have nine claws with a middle body size.

To understand the domestication and genetic basis of the distinct traits of these indigenous chickens, we re-sequenced 25 Daweishan chickens, 10 Hu chickens, 23 Piao chickens, 23 Wuding chickens and four Nine-claw chickens. By comparing the single nucleotide polymorphisms (SNPs) of these indigenous chicken populations with those of 35 RJF individuals as well as of 60 broiler individuals and 56 layer individuals using a rigorous statistic model [6, 51], we were able to find numerous artificial selection signatures in the indigenous chickens, and substantially larger numbers of artificial selection signatures in broilers and layers than previously reported [6, 52, 53]. By comparing the selection signatures between the indigenous chicken breeds, RJF, broilers and layers, we found numerous genomic regions and genes related to the breed-specific traits.

Methods

Re-sequencing short reads from NCBI SRA

We downloaded genomic short reads of two broiler lines from NCBI Sequence Read Archive (SRA): "Broiler A" (*n* = 40, access number PRJEB15276) and "Broiler B" (*n* = 20, access number PRJEB30270). Broiler A was originally from France, and Broiler B was from the company Indian River International. We downloaded DNA short reads of three layer lines from NCBI SRA: "Layer A" (*n* = 25, access number PRJEB15189) were white egg layers, "Layer B" (*n* = 25, access number PRJEB30270) were brown egg layers, and "Layer C" (*n* = 6, access number PRJEB30270) were crossbred layers. We downloaded genomic short reads of two RJF populations from NCBI SRA: "RJF A" (n = 25, access number PRJEB30270) were from northern Thailand, and "RJF B" (*n* = 10, access number PRJEB30270) were from India.

Re-sequencing of indigenous chicken samples

We re-sequenced 85 indigenous chicken individuals from the Experimental Breeding Chicken Farm of the Yunnan Agricultural University (Yunnan, China), including 25 Daweishan chickens aged 10 months (nine males, 16 females), 10 Hu chickens aged 7 months (five males, five females), 23 Piao chickens aged 10 months (11 males, 12 females), 23 Wuding chickens aged 10 months (11 males, 12 females) and four Nine-claw chickens aged 10 months (two males, two females).

Short-reads DNA sequencing

Two milliliters of blood were drawn from the wing vein of each chicken in a centrifuge tube containing anticoagulant (EDTA-2 K) and stored at $-80\,°C$ until use. Genomic DNA (10 μg) in each blood sample was extracted using a DNA extraction kit (DP326, TIANGEN Biotech, Beijing, China) and fragmented using a Bioruptor Pico System (Diagenode, Belgium). DNA fragments around 350 bp were selected using SPRI beads (Beckman Coulter, IN, USA). DNA-sequencing libraries were prepared using Illumina TruSeq® DNA Library Prep Kits (Illumina, CA, USA) following the vendor's instructions. The libraries were subject to 150 cycles paired-end sequencing on an Illumina Novaseq 6000 platform (Illumina, CA, USA) at ~30X coverage.

Real-time quantitative PCR (RT-qPCR) analysis

One to 2 grams of relevant tissues were collected from individual chickens of relevant breeds in a centrifuge

Wu *et al. BMC Genomics*   (2024) 25:428

Page 3 of 20

tube and immediately frozen in liquid nitrogen, then stored at $-80\,°C$ until use. Total RNA from each tissue sample were extracted using TRlzol reagents (TIANGEN Biotech, Beijing China) according to the manufacturer's instructions. RT-qPCR was performed using the Bio-Rad CFX96 real-time PCR platform (Bio-Rad Laboratories. lnc, America) and SYBR Green master mix (iQTM SYBR-Green® Supermix, Dalian TaKaRa Biotechnology Co. Ltd. Add). The primers of the eight genes are listed in Supplementary Note 1. The β-actin gene was used as a reference. Primers were commercially synthesized (Shanghai Shenggong Biochemistry Company P.R.C). Each PCR reaction was performed in 25 μl volumes containing 12.5 μl of iQ™ SYBR Green Supermix, 0.5 μl (10 mM) of each primer, and 1 μl of cDNA. Amplification and detection of products was performed with the following cycle profile: 1 cycle of 95 °C for 2 min, and 40 cycles of 95 °C for 15 s, annealing temperature for 30 s, and 72 °C for 30 s, followed by a final cycle of 72 °C for 10 min. The specificity of the amplification product was verified by electrophoresis on a 0.8% agarose gel and DNA sequencing. The $2^{-\Delta Ct}$ method was used to analyze mRNA abundance. All tissues were analyzed with at least three biological replicates and each biological replicate with five technical replicates, and the means and standard deviation of all these measurements were presented in the relevant figures.

Variant calling

We mapped the short reads of each individual chicken to the reference genome (GRCg7b) using BWA (0.7.17) [54] and SAMtools (1.9) [55] with the default settings and called variants for each individual using GATK-HaplotypeCaller (4.1.6) [56] with the default settings. After generating the GVCF files for each individual, we computed allele frequencies in the same chicken breed using the GATK-CombineGVCFs (4.1.6) tool [56]. For each chicken breed, we removed variants with Quality by depth (QD) < 2, Fisher strand (FS) > 60, Root mean square mapping quality (MQ) < 40, Strand odd ratio (SOR) > 3, Rank Sum Test for mapping qualities (MQRankSum) < − 12.5 and Rank Sum Test for site position within reads (ReadPosRankSum) < − 8 for SNPs and QD < 2, FS > 200, SOR > 10, Likelihood-based test for the consanguinity among samples (InbreedingCoeff) < − 0.8 and ReadPosRankSum < − 20 for indels.

Genetic linkage disequilibrium (LD) analysis

LD analysis in each chicken population was performed using PopLDdecay [57] with the default settings based on the autosome SNPs called in each chicken individual of the 12 chicken populations.

Runs of homozygosity (ROH) analysis

ROH analysis in each chicken population was done using BCFtools (1.10) [58] with the default settings based on the autosome SNPs called in each chicken individual of the 12 chicken populations. For convenience of discussion, we define a LD decay rate as the average distance for LD to decay until $R^2 = 0.2$.

Principal component analysis (PCA)

PCA was performed using PLINK (1.90) [59] with the default settings based on the autosome SNPs called in each chicken individual of the eight breeds.

Population structure analysis

Population structure was inferred using ADMIXTURE (1.3.0) [60] with K = 2, 3, ..., 12 using the default settings based on the autosome SNPs called in each chicken individual of the 12 chicken populations.

Functional annotation of variants

We used the package ANNOVAR [61] to annotate the variants according to their locations on the reference genome into seven categories including 1) intergenic regions, 2) intronic regions, 3) coding regions (synonymous, nonsynonymous, stop gain and stop loss), 4) up/downstream of a gene, 5) splicing sites, 6) 5′ untranslated regions (5'UTRs) and 3′ UTRs, and 7) non-coding RNA regions. We used the tool Ensembl Variant Effect Predictor (VEP) [62] to predict the impact of amino acid-altering SNPs.

Detection of selective sweeps

The selective sweeps were detected using two methods including genetic differentiation (F_{ST}) [63] and nucleotide diversity (π). We estimated F_{ST} for each comparison of two chicken populations using VCFtools (0.1.16) [64] with a sliding window size 40 kb and a step size 20 kb. We estimated π for each group using VCFtools (0.1.16) [64] with a sliding window size 40 kb and a step size 20 kb, and calculated the absolute value of the difference in nucleotide diversity ($|\Delta\pi|$) in each window for each comparison of two populations. For both F_{ST} and $|\Delta\pi|$, we only used the bi-allelic SNPs on autosomes and sex chromosomes for the analysis. To evaluate the statistic significance of the F_{ST} and π values for a comparison, we generate a Null model by shuffling the allele frequency data for 100 times while keeping the SNP positions fixed [51]. We then computed F_{ST} and $|\Delta\pi|$ for the permuted windows as well as their means ($\mu(F_{STNull})$ and $\mu(|\Delta\pi|_{Null})$) and standard deviations ($\sigma(F_{STNull})$ and $\sigma(|\Delta\pi|_{Null})$). We computed

Wu *et al. BMC Genomics*   (2024) 25:428

Page 4 of 20

the Z value for each F_{ST} and $|\Delta\pi|$ values for a comparison by using the following formulas:

$$ZF_{ST}(i) = \left(F_{ST}(i) - \mu(F_{ST\,Null})\right)/\sigma(F_{ST\,Null}) \text{ and}$$

$$Z|\Delta\pi(i)| = \left(|\Delta\pi(i)| - \mu(|\Delta\pi|_{Null})\right)/\sigma(|\Delta\pi|_{Null}).$$

We consider a window with either $ZF_{ST}(i) > 6$ or $Z|\Delta\pi| > 3.09$ (*P*-value < 0.001) to be a putative selective sweep. Since adjacent putative selective sweep windows may overlap with each other, we merged adjacent windows if they overlapped by at least one nucleotide to obtain the discrete selective sweeps (DSSs) for each comparison.

Selective sweeps analysis

To reveal selective sweeps related to different major traits of the domestic chicken populations, we conducted a total of 19 comparisons as summarized in Table 1.

Results

Indigenous chicken breeds have higher nucleotide diversity

Using the re-sequencing short reads of 25 Daweishan chickens, 10 Hu chickens, 23 Piao chickens, 23 Wuding chickens, four Nine-claw chickens, 60 broilers (two populations), 56 layers (three populations) and 35 RJFs (two populations), we detected 20.4, 15.7, 22.0, 19.5 and 13.0 million single nucleotide variants (SNVs) and small indels (< 50 bp) in Daweishan, Hu, Piao, Wuding and Nine-claw chickens (Table 2), respectively. Taking the union of the SNPs found in the different breeds, we ended up with a total of 30 million non-redundant variants in the indigenous chickens (Table 2). Moreover, consistent with the earlier report [52], we found 16.7, 14.4 and 22.3 million SNVs and small indels in the broilers, layers and RJFs (Table 2), respectively, with a total of 27 million non-redundant variants in them. The five indigenous chickens share 20 million (66.7%) variants with the broilers, layers and RJFs. Thus, there are 10 million (33.3%) variants in the five indigenous chickens that are not seen in the broilers, layers and RJFs. Moreover, the broilers and

Table 1 Definitions of the 19 comparisons between different chicken breeds for predicting selective sweeps

Order	Comparisons
1	Daweishan chickens **vs** RJF
2	Hu chickens **vs** RJF
3	Piao chickens **vs** RJF
4	Wuding chickens **vs** RJF
5	Nine-claw chickens **vs** RJF
6	Broilers **vs** RJF
7	Layers **vs** RJF
8	Indigenous chickens (Daweishan chickens + Hu chickens + Piao chickens + Wuding chickens + Nine-claw chickens) **vs** RJF
9	Commercial chickens (layers + broilers) **vs** RJF
10	Indigenous chickens **vs** Commercial chickens
11	Daweishan chickens **vs** the other six domestic chicken breeds
12	Hu chickens **vs** the other six domestic chicken breeds
13	Piao chickens **vs** the other six domestic chicken breeds
14	Wuding chickens **vs** the other six domestic chicken breeds
15	Nine-claw chickens **vs** the other six domestic chicken breeds
16	Broilers **vs** the other six domestic chicken breeds
17	Layers **vs** the other six domestic chicken breeds
18	Daweishan chickens **vs** chickens with relatively large body size (Hu chickens, Wuding chickens and Broilers)
19	Broilers **vs** Layers

Table 2 Summary of genetic variants in the chicken groups

	Daweishan	Hu	Piao	Wuding	Nine-claw	Broiler	Layer	RJF	Indigenous	Commercial
# Individuals	25	10	23	23	4	60	56	35	85	116
# Variants	20,387,469	15,708,834	21,993,907	19,468,846	12,946,552	16,747,786	14,390,952	22,295,197	29,749,839	18,738,449
# Bi-allelic SNPs on autosomes and sex chromosomes	17,267,774	13,446,705	18,646,530	16,484,447	11,225,485	14,162,751	12,096,556	19,053,911	23,405,295	15,692,214
Mean Pi	4.20E-03	3.90E-03	4.40E-03	3.80E-03	4.20E-03	3.40E-03	2.30E-03	3.80E-03	4.30E-03	3.10E-03

Wu *et al. BMC Genomics*   (2024) 25:428

Page 5 of 20

layers have a total of 18.7 million non-redundant variants (Table 2). There are 17.3, 13.5, 18.7, 16.5, 11.2, 14.1, 12.1 and 19.1 million bi-allelic SNPs in the Daweishan, Hu, Piao, Wuding, Nine-claw chickens, broilers, layers and RJFs populations, respectively (Table 2), with a total of 26 million non-redundant bi-allelic SNPs in them. Our subsequent analyses will be focused on these bi-allelic SNPs (Table 2).

We analyzed the mean nucleotide diversity (π) of each chicken breed. Even though both the broilers and the layers were from at least two different sources (Materials and Methods), both had smaller π values than did the five indigenous chickens and RJFs as individual groups or combined groups (Table 2), indicating that commercial chickens have undergone more selection than indigenous chickens, consistent with the breeding histories of commercial and indigenous chickens. The relatively low π values of commercial lines might be due to their close mating and linked selection [52]. Unexpectedly, the mean π value of the indigenous group is higher than that of the RJFs group with two different origins (India and Thailand, Materials and methods, Table 2).

To see the degree of inbreeding of the 12 chicken populations (Materials and Methods), we calculated the average LD for each population as change in average correlation (R^2) between SNPs among all individuals in the population as physical distance increases between SNPs. As shown in Fig. 1a, LD decays much faster in commercial populations than in the indigenous populations, except in nine-claw population where LD decays slower only than in Layer C (crossbred layers), and in

Hu population where LD decays faster than in Broiler B (from Indian River International). Specifically, the average distance for the LD to decay until $R^2 = 0.2$ (LD decay rate) was 197 kb for Layer A (while white egg layer), 177 kb for Layer B (Brown egg layers), 300 kb for Layer C, 78 kb for Boiler A (from France), and 0.6 kb for Broiler B, while it is 300 kb for nine-claw, and only 2.7 kb for Hu, 0.3 kb for Wuding, 0.2 kb for Daweishan and 0.06 kb for Piao. The slow LD decay rate of nine-claw chickens might be due to their small sample size ($n = 4$) and possible sampling bias. Interestingly, although RJF A (from northern Thailand) has a similarly faster LD decay rate (0.07 kb) to those of Piao and Daweishan populations, RJF B (from India) has very slow LD decay rate > 300 kb (Fig. 1a), indicating their possible high degree of inbreeding. Largely consistent with the LD results, the fraction of runs of homozygosity (ROH) is far greater for the autosomes of commercial chickens than indigenous chickens, particularly, for long ROH (> = 500 kb) (Fig. 1b), indicating a higher degree of recent inbreeding of the commercial chickens than indigenous chickens. Moreover, RJF B has high fraction of long ROH similar to that of commercial chicken populations (Fig. 1b), further indicating that RJF B has a high degree of inbreeding.

Variants are enriched in non-coding regions while those in coding regions are largely tolerant

Based on the location of the bi-allelic SNPs, we classified them into seven categories, including intergenic (variants in intergenic regions), intronic (variants in introns), up/downstream (variants within a 1 kb region
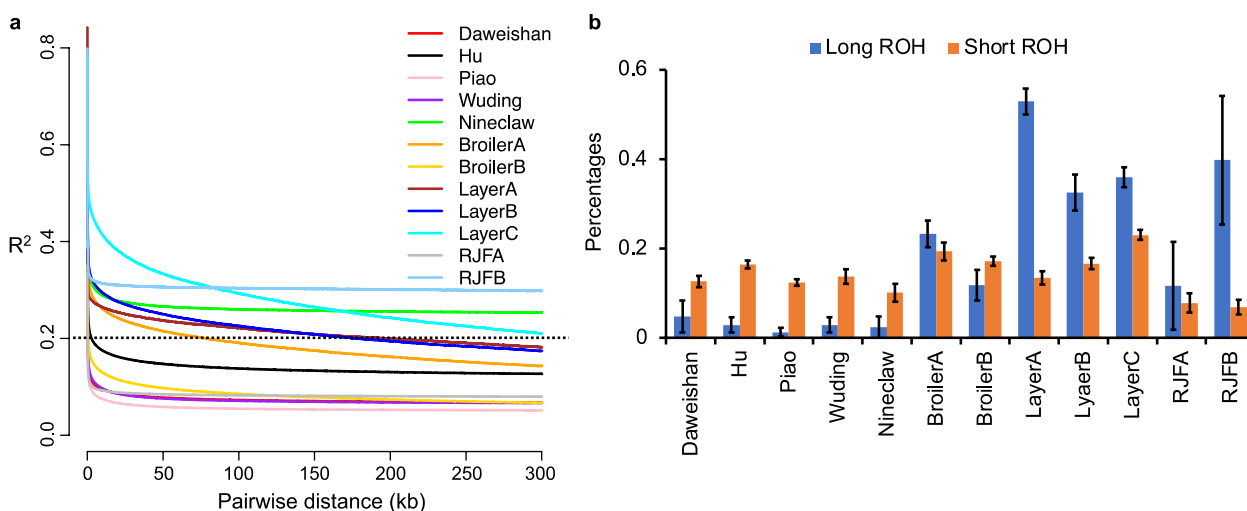

Fig. 1 Genetic linkage disequilibrium (LD) and runs of homozygosity (ROHs) in 12 chicken populations. **a.** LD decay in each chicken population as change in average correlation (R^2) between SNPs among all individuals per population as physical distance increases between SNPs. The dashed horizontal line indicates the $R^2 = 0.2$ line to define the LD decay rate (Materials and Methods). **b.** Percentages of ROH on autosomes. Blue bars show fractions of long ROH ≥0.5 Mb, and yellow bars fractions of short ROH < 0.5 Mb. Error bars are SD

Wu *et al. BMC Genomics*   (2024) 25:428

Page 6 of 20

up/downstream of transcription start/end sites), splicing (variants within 2 bp of a splicing junction), UTR 3′/ UTR 5′ (variants in 3′/5′ untranslated regions), ncRNA (variants in non-coding RNA genes) and coding (variants in coding sequences). The relative abundances of the bi-allelic SNPs in each chicken group are shown in Table S1. Specifically, of the bi-allelic SNPs in each chicken group, 29.71% ~ 30.46% fall within intergenic regions, 50.44% ~ 51.57% are located in intronic regions, 2.86% ~ 2.97% fall within up/downstream regions, 4.08% ~ 4.24% are located in 3′ UTR/5′ UTR regions, 0.01% fall within splicing regions, 10.09 ~ 10.18% are located in ncRNA regions, and 1.53% ~ 1.80% fall within coding regions (Table S1). Therefore, only a small portion (1.53% ~ 1.80%) of the bi-allelic SNPs fall within coding regions, while the vast majority (98.20% ~ 98.47%) are located in non-coding regions. As non-coding regions comprise 96.92% of the reference chicken genome (GRCg7b assembly), the SNPs are enriched in non-coding regions relative to in coding regions.

Among the bi-allelic SNPs in coding regions of each chicken group, 33.33% ~ 38.33% (Daweishan = 35.71%, Hu = 34.55%, Piao = 36.47%, Wuding = 35.71%, Nine-claw = 33.33%, broiler = 34.97%, layer = 35.22%, RJF = 35.80%, indigenous = 38.33%, commercial = 36.36%) are amino acid-altering (AA-altering, i.e., nonsynonymous and stop-gain/loss) SNPs (Table S1). Among the nonsynonymous SNPs in each chicken group, most (Daweishan = 86.44%, Hu = 87.50%, Piao = 85.24%, Wuding = 86.44%, Nine-claw = 88.00%, broiler = 87.50%, layer = 87.27%, RJF = 85.96%, indigenous = 63.24%, commercial = 69.49%) are tolerant SNPs and only a small proportion (Daweishan = 13.56%, Hu = 12.50%, Piao = 14.76%, Wuding = 13.56%, Nine-claw = 12.00%, broiler = 12.50%, layer = 12.73%, RJF = 14.04%, indigenous = 36.76%, commercial = 30.51%) are intolerant, which might be deleterious variants and thus under purifying selection in the chicken group (Table S1).

Indigenous chicken breeds have a high portion of rare nonsynonymous SNPs

We compared the allele frequencies of the SNPs in coding regions in the groups of RJFs, indigenous and commercial populations. As shown in Fig. 2a, all the three groups have higher portion of rare (< 0.005) nonsynonymous SNPs than rare synonymous SNPs, but the opposite is true for more frequent (> 0.01) ones, indicating that rare nonsynonymous SNPs tend to be deleterious and thus under purifying selection. The same conclusion has been drawn in an earlier study for commercial chickens [52]. Interestingly, the indigenous chickens have the highest rare allele frequency densities for both nonsynonymous and synonymous SNPs, followed by

RJFs and the commercial chickens. The earlier study also noted that RJFs had higher rare allele frequency densities than commercial chickens [52]. Among the five different indigenous chicken breeds and two commercial chicken breeds, Piao chickens and the layers have the highest and the lowest rare allele densities (Fig. S1), consistent with their highest and lowest π values, respectively (Table 2).

Only a small portion of breed-specific SNPs are fixed

We analyzed the group-specific SNPs in each chicken group, and found that RJFs have the highest number of unique SNPs (2.9 million) among the eight chicken groups (Table S2), which is consistent with the finding in the previous study [52], suggesting a loss of ancestral alleles during chicken domestication. Except for Hu chicken and Nine-claw chicken with a small population size (Table 2), layers have the lowest number of unique SNPs (455 k) and broilers have the second lowest number of unique SNPs (520 k) among the eight chicken breeds (Table S2), while Daweishan, Piao and Wuding chickens have 1.1, 1.3 and 0.7 million unique SNPs, respectively, indicating once again that genetic diversity of indigenous chickens is higher than those of the layers and the broilers. From 0.83% (RJFs) to 1.39% (Wuding chicken) of the group-specific SNPs are missense mutations (Table S2). Most of the group-specific SNPs have allele frequencies lower than 0.5, and only a very small portion (0.05% ~ 0.59%) are close to being fixed (allele frequency > 0.9) in all the eight groups of chickens (Table S2). The same is true for the missense SNPs (0.04% ~ 0.48%) (Table S2). More details of the group-specific missense SNPs and affected genes are listed in Tables S3–S10.

We also compared allele frequency spectrums of the group-specific SNPs in the eight chicken groups (Fig. 2b). Group-specific alleles of the layers tend to have higher frequencies than those in other groups (except for Hu chicken and Nine-claw chicken with a small population size), consistent with the finding in the earlier study [52].

The indigenous chickens are more closely related to one another while commercial chickens are genetically different

To reveal the genetic relationships of the individual chickens, we performed a PCA based on occurring patterns of the 26 million bi-allelic SNPs. As shown in Fig. 2c, the five indigenous populations from Yunnan Province, China, are clustered together with the RJF A population from northern Thailand that is geographically close to Yunnan Province, China, while the RJF B population from India form a separate cluster nearby. On the other hand, Layer B and Broiler B populations form two tight clusters close to the cluster of indigenous chickens and
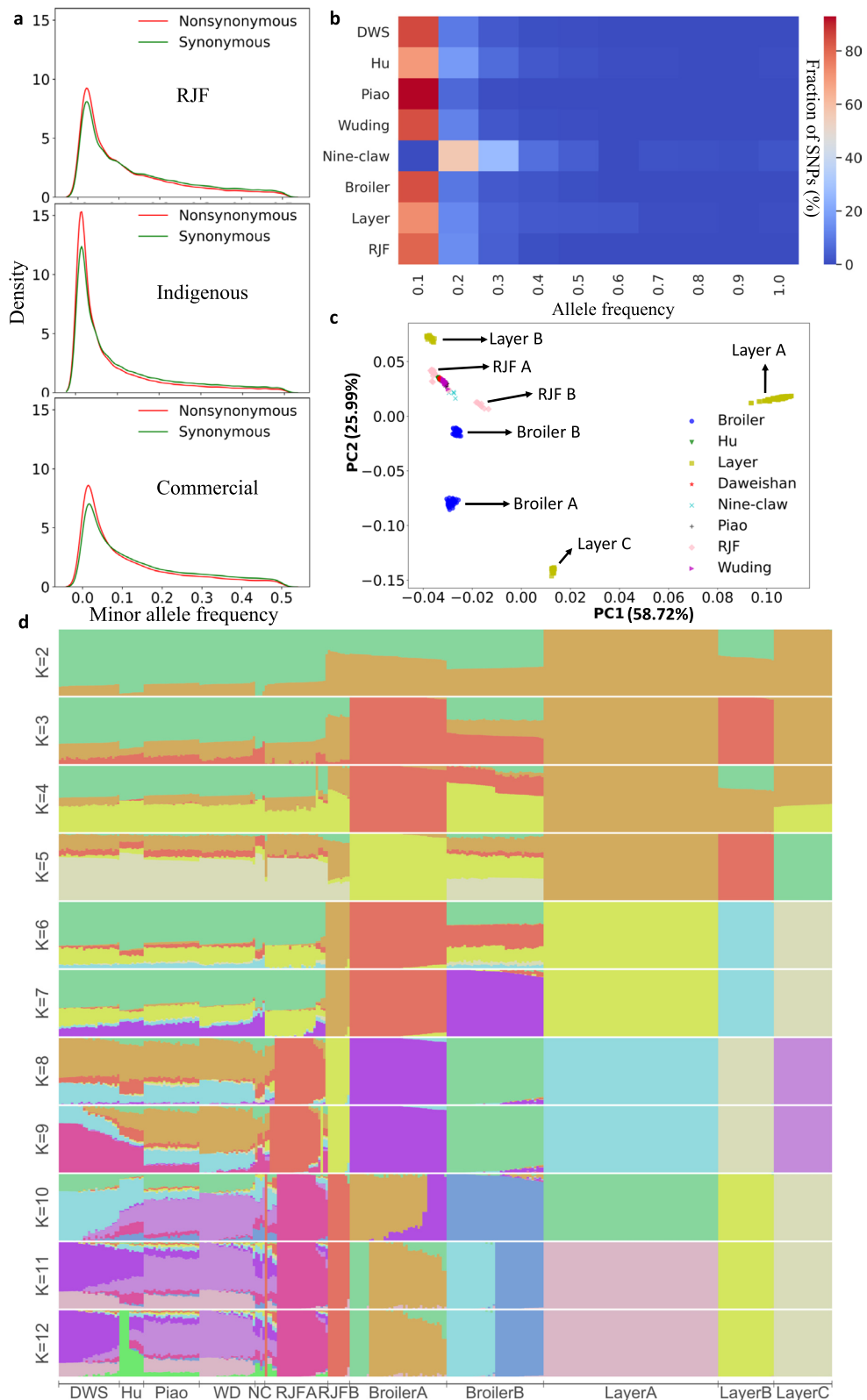
Fig. 2 Analysis of frequency spectrums of SNPs. **a.** Distribution of minor allele frequency of synonymous and nonsynonymous SNPs in different chicken groups. **b.** Heatmap of allele frequency of group-specific SNPs. **c.** Principal component analysis of chicken population based on the detected 26 million SNPs. **d.** Genetic structure of the 12 chicken populations estimated using the ADMIXTURE program for K = 2, 3, ..., 12

Wu *et al. BMC Genomics*   (2024) 25:428

Page 8 of 20

RJF A populations, while Layer A, Layer C and Broiler A populations form three distinct clusters far away from the cluster of indigenous chickens and RJF A populations.

We also analyzed genetic structure of the 12 chicken populations based on the 26 million bi-allelic SNPs. As shown in Fig. 2d, when K=7, 8 or 9, Broiler A, Broiler B, Layer A, Layer B, Layer C, RJF A and RJF B populations form distinct clusters with little admixture among them, indicating that each of these populations have distinct genetic variations. However, when K=11 and 12, both Broiler A and Broiler B populations are further clustered in two distinct subclusters, with small admixture of a Broiler A subcluster in the other Broiler A subcluster (Fig. 2d), indicating that each of the two broiler populations are actually made up of two distinct subpopulations. In contrast, none of the four indigenous chicken populations form a distinct cluster for all K > 2 evaluated (Fig. 2d). Instead, all the five the indigenous populations have similar genetic admixture and share small portions of genetic variations with RJF A (Fig. 2d). Both PCA and admixture analyses indicate that the five indigenous breeds are closely related to one another, and they are also more closely related to RJF A from northern Thailand than to RJF B from India, while commercial chickens are genetically different even for those with similar productivities.

A rigorous null model facilitates sensitive detection of selective sweeps

To detect selection signatures of each chicken breed, we identified selective sweeps [65] along each chromosome based on the frequencies of the bi-allelic SNPs using both the F_{ST} and π statistics. We found (see below) that ZF_{ST} is more sensitive than $Z|\Delta\pi|$ to identify selective sweeps,

thus a higher cutoff of ZF_{ST} (6 vs 3) is used here. Since adjacent windows can overlap with each other, we merged the overlapping selective sweep windows to form a discrete selective sweep (DSS) (Materials and methods). To find selection signatures of the chicken groups, we conducted a total of 19 different comparisons (Tables 3 and S11). The selective sweep windows identified by either of the two methods are distributed along all the chromosomes with varying densities (Figs. 3, 4, S2 and S3). We generally detected more DSSs using ZF_{ST} (806~2125 DSSs) than using $Z|\Delta\pi|$ (110~818 DSSs) for all the 19 comparisons (Tables 3 and S11) even using a higher ZF_{ST} cutoff. However, less than 60% (9.16%~58.23%) of the DSSs identified by $Z|\Delta\pi|$ overlap with those identified by ZF_{ST} (Tables S12~S30) in the 19 comparisons, indicating that the results of the two methods are largely complementary with each other. We thus take their union as the final prediction of DSSs (Tables S12–S30). We finally identified 1073~2745 DSSs consisting of 1998~7284 windows containing 528~2147 genes in each of the 19 comparisons (Tables 3 and S11). Therefore, we find much more selective sweeps than the previous study using a mixture model (~60 selective sweep windows of 40kb size) [52]. The DSSs in the 19 comparisons have a varying length ranging from 40 kbp to 2240 kbp with a median length of 60 kbp, and 91.71% of them are shorter than 140 kbp (Fig. S4a). The total length of the DSSs in each comparison consist of 5.85% (Nine Claw vs RJF)~18.88% (Broiler vs Layer) of the reference genome (GRCg7b assembly) (Fig. S4b). In general, comparisons with broilers alone as one of the two compared groups tend to have a high portion of genome under selection (Fig. S4b), suggesting that broilers have gone through most extensive selection.

Table 3 Summary of putative selective sweeps and DSSs found in each chicken group in comparison with the RJFs

Breeds	# Analyzed windows		# Windows passed critical cut-off		# Discrete selective sweeps (# genes)		# Final putative selective sweep windows (# DSSs, # genes in DSSs)
	ZF_{ST}	$Z|\Delta Pi|$	$ZF_{ST} > 6$	$Z|\Delta Pi| > 3.09$	$ZF_{ST} > 6$	$Z|\Delta Pi| > 3.09$	
Daweishan VS. RJF	51,857	51,789	4,157	595	1,268 (978)	263 (790)	4,619 (1,550, 1706)
Hu VS. RJF	51,822	51,719	3,407	1,601	1,054 (783)	645 (802)	4,415 (1,811, 1,425)
Piao VS. RJF	51,860	51,789	4,057	383	1,313 (845)	152 (647)	4,381 (1,483, 1,455)
Wuding VS. RJF	51,853	51,782	3,752	1,554	1,255 (811)	764 (892)	4,962 (2,077, 1,617)
Nine-claw VS. RJF	51,769	51,615	1,640	497	806 (398)	262 (176)	1,998 (1,073, 528)
Broiler VS. RJF	51,643	51,756	6,366	2,150	1,784 (1,477)	818 (646)	7,012 (2,745, 1,661)
Layer VS. RJF	51,808	51,725	3,808	791	1,521 (896)	407 (404)	4,192 (1,926, 1,139)
Indigenous VS. RJF	51,886	51,793	4,294	301	1,183 (817)	130 (528)	4,572 (1,319, 1,333)
Commercial VS. RJF	51,658	51,767	5,063	1,146	1,787 (1,191)	544 (584)	5,540 (2,382, 1,544)

(See figure on next page.)

Fig. 3 Manhattan plots of ZF_{ST} values of each window on each chromosome for the indicated comparisons. The blue horizontal line indicates the ZF_{ST} cutoff = 6. Examples of genes in significant selective sweep windows are shown in different color. Genes that have been previously reported in selective sweep windows are shown in red, genes in our predicted selective sweep windows potentially related to the specific traits of each chicken breed are shown in blue, and genes in novel selective sweep windows with extremely high ZF_{ST} values are shown in black. Asterisk represents selective sweep windows lacking annotated genes
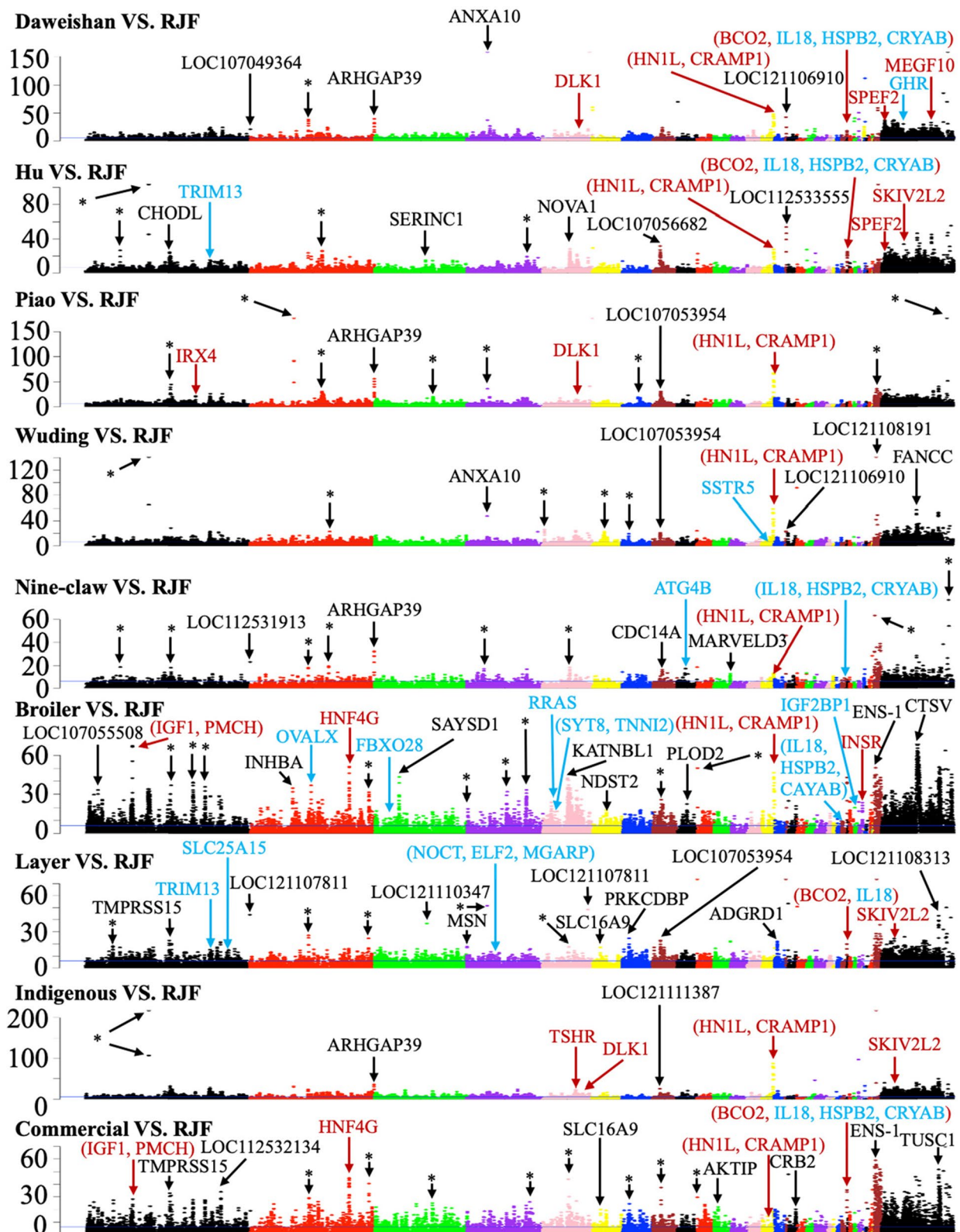
Fig. 3 (See legend on previous page.)

The 19 comparisons reveal selection signatures of the chicken groups

Firstly, to find the genetic differences between artificial selection and natural selection, we compared each chicken breed (including indigenous chicken group and commercial chicken group) with the RJFs. As summarized in Table 3 and Tables S12～S20, we identified varying numbers of selective sweeps (1998～7012) and DSSs (1073～2745) involving 528～1706 genes for the eight comparisons, suggesting that these chicken breeds might have gone through different levels of artificial selection. For example, the Broiler vs RJF comparison yields the highest numbers of selective sweeps (7012) and DSSs (2745), suggesting again that broilers might have gone through the most intensive artificial selection. Among the five indigenous breeds, Wuding chickens might have gone through the most intensive artificial selection with the highest numbers of selective sweeps (4962) and DSSs (2077). In addition, we identified more selective sweeps (5540) and DSSs (2382) for the Commercial vs RJF comparison than those (4572 selective sweeps and 1319 DSSs) in Indigenous vs RJF comparison (Table 3), suggesting that the commercial chickens have gone through more intensive artificial selections than indigenous chickens as generally understood. Interestingly, hundreds of affected genes in putative selective sweeps in each indigenous breed were also found in putative selective sweeps in broilers and layers (Fig. S5).

Secondly, to reveal the genetic differences between traditional selection and industrial selection, we compared the indigenous chicken group with the commercial chicken group and identified a large number of selective sweeps (6735) and DSSs (2532) involving 2147 genes (Tables S11 and S21). This result suggests that indigenous chickens and the commercial chickens have gone through quite different artificial selection routes as commonly understood.

Thirdly, to reveal unique selective sweeps of each chicken breed, we compared each domestic chicken breed with the rest domestic chicken breeds and found that broilers and layers have much higher number of unique selective sweeps (7239 and 6401, respectively) and DSSs (2514 and 2479, respectively) than the indigenous breeds (2801～4555 and 1363～1890, respectively) (Tables S11 and S22～28).

Fourthly, to reveal possible selective sweeps underlying the miniature body size of Daweishan chicken, we compared Daweishan chicken with the group of Hu chicken, Wuding chicken and Broiler (HWB), with a relatively large body size, and identified 6359 selective sweeps and 2263 DSSs including 1911 genes (Tables S11 and S29).

Finally, to find the selection difference between broilers and layers, we compared the two groups and identified 7284 selective sweeps and 2630 DSSs including 1802 genes (Tables S11 and S30). For the similar comparison in a previous study [52], only 41 selective sweeps (40 kb) were identified using a mixture model for normalizing F_{ST} and $\Delta\pi$. Therefore, we identified substantially more selective sweeps by employing a more rigorous Null model.

Amino-acid altering SNPs are enriched in the selective sweeps

To identify selective sweeps that might be responsible for the formations of the indigenous chickens, the broiders and the layers, we took the union of DSSs found in comparisons with a breed alone as one of the compared group, e.g., for Daweishan chicken, we took the union of DSSs in comparisons Daweishan vs RJF, Daweishan vs other-breeds and Daweishan vs HWB; and for Hu chicken, we took the union of DSSs in comparisons Hu vs RJF and Hu vs other-breeds; and so on. We identified from 1.2 million (Nine-claw chicken) to 4.1 million (broilers) SNPs in the union of DSSs in each domestic chicken breed (Table S31). Among these SNPs, only 1.3% (Nine-claw chicken)～2.0% (Hu chicken) are located in coding regions, while the remaining vast majority (97.98%～98.72%) fall in non-coding regions (Table S31). As non-coding regions comprise 96.92% of the reference chicken genome (GRCg7b assembly), as in the case of all the bi-allelic SNPs (Table S1), the SNPs in the DSSs are also enriched in non-coding regions relative to in coding regions. Among the SNPs in coding regions, 36.30%～45.23% are amino-acid altering, which are higher than the corresponding values of all the bi-allelic SNPs (33.33%～38.33%) (Table S1), suggesting that amino-acid altering SNPs are enriched in the selective sweeps relative to all the bi-allelic SNPs ($p = 0.005$, K-S test).

(See figure on next page.)

Fig. 4 Manhattan plots of ZⅠΔπⅠ values of each window on each chromosome for the indicated comparisons. The blue horizontal line indicates the ZⅠΔπⅠ cutoff = 3.09. Examples of genes in significant selective sweep windows are shown in different color. Genes that have been previously reported in selective sweep windows are shown in red, genes in our predicted selective sweep windows potentially related to the specific traits of each chicken breed are shown in blue, and genes in novel selective sweep windows with extremely high ZⅠΔπⅠ values are shown in black. Asterisk represents selective sweep windows lacking annotated genes
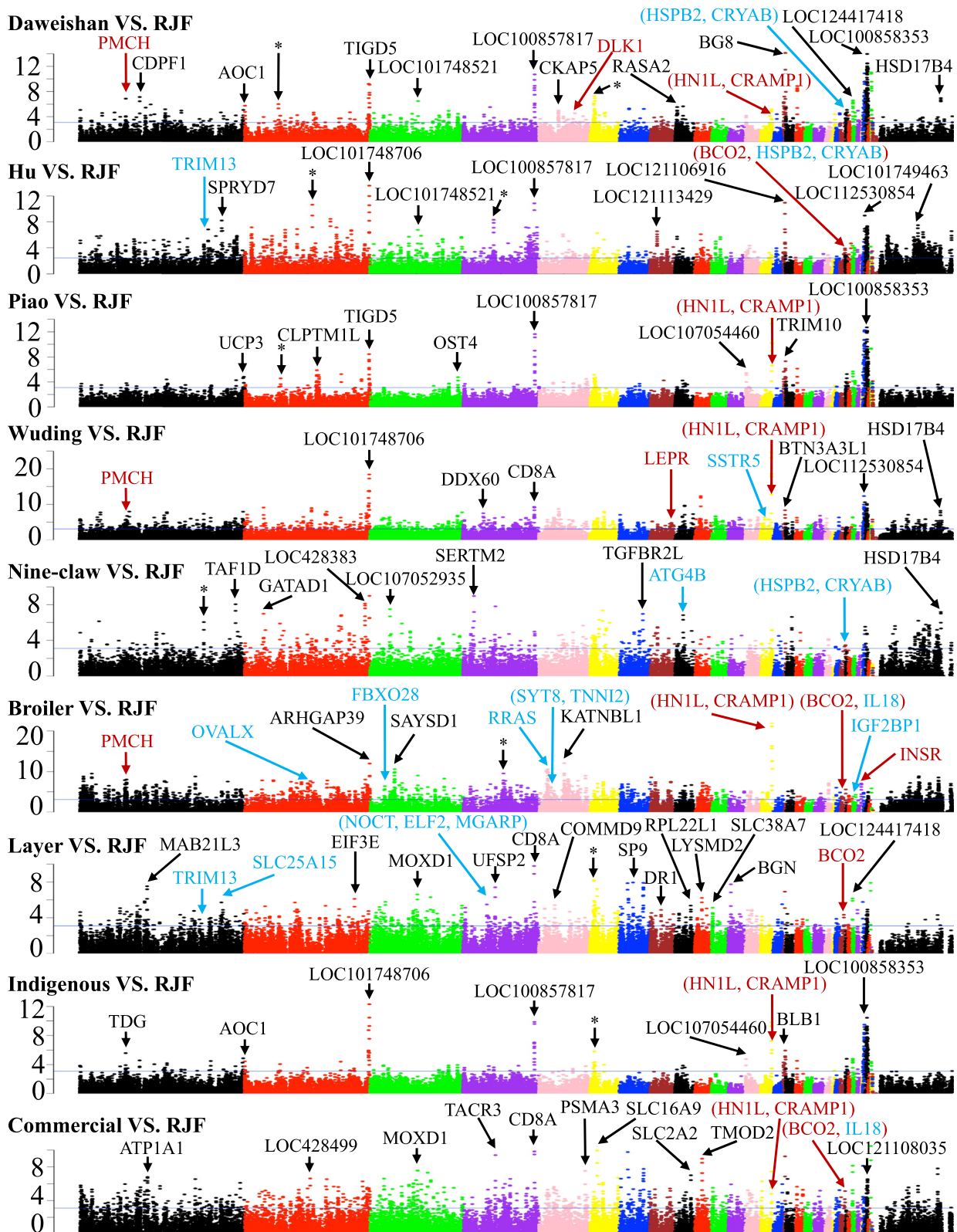
Fig. 4 (See legend on previous page.)

Our predicted selective sweeps are supported by experimental data

To evaluate our detected selective sweeps, we first compared them (Tables S12~S30) with the 15,439 QTLs in the chicken QTL database [66]. We find that 90.5%~98.3% putative DSSs in each of our comparisons overlap one or more QTLs in the chicken QTLdb (Tables 4 and S32). On the other hand, we find that 23.9%~41.6% QTLs in chicken QTLdb overlap one or more our predicted DSSs in each of our comparisons (Tables 4 and S32), and 11,449 (74.2%) QTLs in the chicken QTLdb overlap one or more of our predicted DSSs in different comparisons.

As an additional validation of our detected selective sweeps, we next compared the genes in our predicted DSSs with those that have been reported to be under selection during chicken domestication process, and we describe a few examples here. 1) It has been shown that the BCO2 locus is involved in the yellow skin trait in domestic chickens [67], and we confirmed this results in several of our comparisons, including Daweishan vs RJF (ZF_{ST} = 11.27), Daweishan vs HWB (ZF_{ST} = 14.79), Hu vs RJF (ZF_{ST} = 18.26, $Z|\Delta\pi|$ = 4.86), Hu vs other-breeds (ZF_{ST} = 6.27), Piao vs other-breeds (ZF_{ST} = 29.36), Wuding vs other-breeds (ZF_{ST} = 22.18), Broiler vs RJF ($Z|\Delta\pi|$ = 5.08), Broiler vs Layer (ZF_{ST} = 9.90), Layer vs RJF (ZF_{ST} = 15.66, $Z|\Delta\pi|$ = 3.17), Layer vs other-breeds (ZF_{ST} = 8.70), Indigenous vs Commercial (ZF_{ST} = 26.75, $Z|\Delta\pi|$ = 3.67) and Commercial vs RJF (ZF_{ST} = 27.16, $Z|\Delta\pi|$ = 4.44) (Figs. 3, 4, S2 and S3); 2) The *TSHR* locus is known to be involved in regulation of reproduction and metabolic functions in commercial chickens [53], and we found that the locus was in selective sweeps in comparisons Hu vs other-breeds (ZF_{ST} = 11.75) and Indigenous vs RJF (ZF_{ST} = 13.43) (Figs. 3 and S2); 3) It has been reported that the *HNF4G* and *IGF1* loci are involved in growth regulation in chicken [53, 68], and we found that the two loci were in selective sweeps in comparisons Broiler vs RJF (for HNF4G, ZF_{ST} = 24.27; for IGF1, ZF_{ST} = 32.54), Commercial vs RJF (for *HNF4G*, ZF_{ST} = 26.25; for *IGF1*, ZF_{ST} = 15.51) and Indigenous vs Commercial (for *HNF4G*, ZF_{ST} = 24.51; for *IGF1*, ZF_{ST} = 17.86)

(Figs. 3 and S2); 4) It has been reported that the *PMCH* locus is related to regulation of appetite and metabolic functions [53, 69], and we found that the locus was in the selective sweeps in several of our comparisons including Daweishan vs RJF ($Z|\Delta\pi|$ = 3.25), Daweishan vs HWB ($Z|\Delta\pi|$ = 6.04), Wuding vs RJF ($Z|\Delta\pi|$ = 4.82), Wuding vs other-breeds ($Z|\Delta\pi|$ = 4.07), Broiler vs RJF (ZF_{ST} = 32.54, $Z|\Delta\pi|$ = 6.05), Broiler vs other-breeds (ZF_{ST} = 19.59, $Z|\Delta\pi|$ = 6.34), Broiler vs Layer (ZF_{ST} = 27.21), Commercial vs RJF (ZF_{ST} = 15.51) and Indigenous vs Commercial (F_{ST} = 17.86) (Figs. 3, 4, S2 and S3); 5) It has been shown that the INSR locus is related to the growth of chicken by encoding a critical component in insulin signaling [53], and we found that the locus was in the selective sweeps in comparisons Daweishan vs HWB (ZF_{ST} = 7.37, $Z|\Delta\pi|$ = 4.46), Broiler vs RJF (ZF_{ST} = 15.30, $Z|\Delta\pi|$ = 5.23), Broiler vs other-breeds (ZF_{ST} = 11.61, $Z|\Delta\pi|$ = 4.86) and Broiler vs Layer (ZF_{ST} = 9.78) (Figs. 3, 4, S2 and S3); 6) It has been shown that the *NELL*1 locus is related to the skeletal integrity in broiler [68], and we found that the locus was in the selective sweeps of the Broiler vs other-breeds comparison (ZF_{ST} = 22.06) (Fig. S2); 7) It has been reported that the IRX4 locus is related to the rumpless trait of Piao chicken [45], and we found that the locus was in the selective sweeps in comparisons Piao vs RJF (ZF_{ST} = 13.79) and Piao vs other-breeds (ZF_{ST} = 12.12) (Figs. 3 and S2). The other selective sweep loci found in the previous studies are also confirmed by our results, such as *ALX*1, *KITLG*, *EGFR*, *DLK*1, *JPT*2 (annotated as *HN1L* in GRCg7b), *CRAMP*1 and *GLI*3 loci, which are related to the general domestication process of chicken [52], the *SKIV2L*2 locus that is related to pigmentation [52], and the *LEPR*, *MEGF*10 and *SPEF*2 loci, which are related to production-oriented selection [52] (Figs. 3, 4, S2 and S3).

Novel selective sweeps are found in the chicken breeds

In addition to confirming many previously identified selective sweeps containing genes related to chicken domestication as described above, we also find numerous novel selective sweeps containing genes (Tables S12~S30) or in gene deserts. We now highlight a

Table 4 Summary of putative DSSs overlapped with chicken QTLs

Breeds	# Putative DSSs	# Putative DSSs overlap QTLs in chicken QTLdb (%)	# QTLs in chicken QTLdb overlap our putative DSSs (%)
Daweishan VS. RJF	1,550	1,472 (95.0%)	5,526 (35.8%)
Hu VS. RJF	1,811	1,702 (94.0%)	5,247 (34.0%)
Piao VS. RJF	1,483	1,342 (90.5%)	4,801 (31.1%)
Wuding VS. RJF	2,077	1,979 (95.3%)	5,381 (34.9%)
Nine-claw VS. RJF	1,073	990 (92.3%)	3,686 (23.9%)
Broiler VS. RJF	2,745	2,601 (94.8%)	5,463 (35.4%)
Layer VS. RJF	1,926	1,837 (95.4%)	5,334 (34.5%)
Indigenous VS. RJF	1,319	1,266 (96.0%)	5,010 (32.5%)
Commercial VS. RJF	2,382	2,237 (93.9%)	5,086 (32.9%)

few of them with extremely high ZF$_{ST}$ and/or Z|Δπ| values in each comparison (Figs. 3, 4, S2 and S3). 1) Gene ARHGAP39 on chromosome 2 is in the selective sweeps with extremely high ZF$_{ST}$ and/or Z|Δπ| values in comparisons Daweishan vs RJF, Piao vs RJF, Nine-claw vs RJF, Indigenous vs RJF, Nine-claw vs other-breeds and Broiler vs RJF. ARHGAP39 plays important roles in cell cytoskeletal organization, growth, differentiation, neuronal development and synaptic functions [70]; 2) Gene TIGD5 on chromosome 2 is in the selective sweeps with extremely high Z|Δπ| values in comparisons Daweishan vs RJF, Piao vs RJF, Daweishan vs other-breeds, Hu vs other-breeds, Piao vs other-breeds, Wuding vs other-breeds, Daweishan vs HWB and Indigenous vs Commercial. *TIGD*5 encodes the tigger transposable element-derived protein 5 and is important for nucleic acid binding [71]; 3) Gene *KCNK*16 on chromosome 3 is in the selective sweeps with extremely high ZF$_{ST}$ values in comparisons Layer vs other-breeds and Broiler vs Layer. *KCNK*16 encodes a rapidly activating and non-inactivating outward rectifier K$^+$ channel [72]; 4) Gene *CD8A* on chromosome 4 is in the selective sweeps with extremely high Z|Δπ| values in comparisons Wuding vs RJF, Layer vs RJF, Commercial vs RJF, Daweishan vs other-breeds, Piao vs other-breeds, Wuding vs other-breeds, Nine-claw vs other-breeds, Broiler vs other-breeds, Layer vs other-breeds, Indigenous vs Commercial and Daweishan vs HWB. CD8A encodes the T-cell surface glycoprotein CD8 alpha chain precursor and plays essential roles in immune response [73]; 5) Gene *COL6A*2 on chromosome 7 is in the selective sweeps with extremely high ZF$_{ST}$ values in comparisons Daweishan vs other-breeds, Daweishan vs HWB and Piao vs other-breeds. The gene encodes the collagen alpha-2(VI) chain precursor which act as a cell-binding protein [74]. Besides the genes mentioned above, we also indicate in Figs. 3, 4, S2 and S3 many other genes located in novel selective sweeps with extremely high ZF$_{ST}$ and/or Z|Δπ| values in multiple comparisons such as: *ANXA*10 on chromosome 3, gene *LOC107053954* on chromosome 8, gene *SLC16A*9 on chromosome 6, gene *ENS*-1 on chromosome W and gene *HSD17B*4 on chromosome Z. It is interesting to experimentally investigate the roles of these genes in the domestication and breeding of each chicken breed.

In Figs. 3, 4, S2 and S3, we also label a few examples of selective sweeps in gene desserts, with extremely high ZF$_{ST}$ and/or Z|Δπ| values in multiple comparisons. It is highly likely that these selective sweeps might harbor non-coding functional sequences such as cis-regulatory modules of distal genes.

Selective sweeps related to each chicken breed

In addition to finding numerous novel selective sweeps containing genes in each comparison (Tables S12∼S30), we also identify numerous unique selective sweeps/DSSs that are only seen in comparisons with a breed alone as one of the two compared groups or selective sweeps/DSSs containing genes with interesting functions. These selective sweeps/DSSs might contain genes (Tables S33∼S39) whose functions are related to the specific traits of the chicken breed. Specifically, for Daweishan chicken, we identified 44 putative genes in the selective sweeps that might be related to its unique traits including the small body size (Table S33). For example, the *GHR* (growth hormone receptor) gene is located in a selective sweep window on chromosome Z, which overlaps body weight QTLs and shank length QTLs. The gene is in the selective sweep windows identified in comparisons with Daweishan chicken alone as one of the two compared groups (Daweishan vs RJF, ZF$_{ST}$ = 17.71; Daweishan vs other-breeds, ZF$_{ST}$ = 17.69; Daweishan vs HWB, ZF$_{ST}$ = 19.51) (Figs. 3 and S2). It has been reported that loss-of-function mutations in GHR was related to sex-linked dwarfism in chicken [75]. We analyzed the SNPs in the GHR gene body for each chicken breeds using the GRCg7b assembly as the template and found 79 unique SNPs in the gene of Daweishan chicken, which were not present in the other chicken breeds (Hu, Piao, Wuding, Nine-claw, Broiler, Layer and RJF). Among these 79 unique SNPs, 68 are in intronic regions, 10 are in UTRs and one is nonsynonymous SNP that leads to a CGG to TGG (R to W) codon mutation, which is predicted to be tolerant by VEP [62]. The substitution allele has a frequency of 0.76, thus it is only nearly fixed. As no fixed potential amino acid-altering mutation could be found in the GHR coding regions, we hypothesize that *GHR* gene might be related to the small body size of Daweishan chicken through changes in its regulatory sequences in the window, resulting in changes in its expression.

For Hu chicken, we identified 14 putative genes in selective sweeps (Table S34) that might be related to its unique traits including the very stout legs. Specifically, gene *TRIM*13 in a selective sweep on chromosome 1 (Hu vs RJF, ZF$_{ST}$ = 7.49 and Z|Δπ| = 7.19; Hu vs other-breeds, ZF$_{ST}$ = 10.01 and Z|Δπ| = 4.58) (Figs. 3, 4, S2 and S3) overlaps shank circumference QTLs, and there are two nonsynonymous SNPs in the gene body which are fixed (allele frequency=1). Thus, it is interesting to experimentally investigate the role of *TRIM*13 in the development of the very stout legs of Hu chicken.

For Piao chicken, we identified six putative genes in selective sweeps (Table S35) that might be related its unique traits including the rumpless trait. Of these

Wu *et al. BMC Genomics*   (2024) 25:428

Page 14 of 20

six genes, *IRX*4 in a selective sweep on chromosome 2 was reported to be related to the rumpless trait of Piao chicken in a previous study [45], and we also found that the selective sweeps were only identified by our comparisons Piao vs RJF (ZF_{ST} = 13.79) and Piao vs other-breeds (ZF_{ST} = 12.12) among the 19 comparisons (Figs. 3 and S2, Tables S14 and S24). Thus, it is highly likely that IRX4 is related to the rumpless trait of Piao chicken. At the same time, the previous study also identified genes *IL*18, *HSPB*2, and *CRYAB* to be related to the rumpless trait of Piao chicken. Although we also found these three genes in the selective sweeps of the comparison Piao vs other-breeds (Table S24 and Fig. S2), these three genes were also present in the selective sweeps of comparisons with a breed having normal tails alone as one of the two compared groups, such as Daweishan vs RJF (Table S12, Figs. 3 and S2), Hu vs RJF (Table S13, Figs. 3 and S2), Nine-claw vs RJF (Table S16, Figs. 3 and S2), Broiler vs RJF (Table S17, Figs. 3 and S2), Layer vs RJF (Table S18, Fig. 3) and Daweishan vs HWB (Table S29, Fig. 3). Therefore, these three genes might not be related to the rumpless trait of Piao chicken.

For Wuding chicken, we identified 18 putative genes in selective sweeps (Table S36) that might be related to its unique traits including colorful feathers and thick fat. Specifically, gene SSTR5 in a selective sweep on chromosome 14 (Wuding vs RJF, ZF_{ST} = 7.38 and $Z|\Delta\pi|$ = 5.16; Wuding vs other-breeds, $Z|\Delta\pi|$ = 3.33) (Figs. 3, 4 and S3) overlaps body weight QTLs, however, there are no nonsynonymous SNPs in its gene body. Gene *LOC101748311* in a selective sweep on chromosome 1 (Wuding vs other-breeds comparison, ZF_{ST} = 9.70 and $Z|\Delta\pi|$ = 3.59) (Figs. S2 and S3) overlaps the feather density QTLs and comb length QTLs and there are two nonsynonymous SNPs in its gene body, but their allele frequencies are very low (< 0.2). It is likely that both genes might be related to Wuding chicken's traits by changes in regulatory regions, which warrants further experimental studies.

For Nine-claw chicken, we identified seven putative genes in selective sweeps (Table S37) that might be related to its unique traits. Specifically, gene *ATG4B* on chromosome 9 (Nine-claw vs RJF, ZF_{ST} = 7.00 and $Z|\Delta\pi|$ = 3.60; Nine-claw vs other-breeds, ZF_{ST} = 7.56) (Figs. 3, 4 and S2) overlaps egg production rate QTLs, but there are no nonsynonymous SNPs in its gene body.

For Broilers, we identified 151 putative genes in selective sweeps (Table S38) that might be related to its unique traits including the fast growth rate. Of these genes, *GHRHR* on chromosome 2 (Growth hormone-releasing hormone receptor) (Broiler vs other-breeds, ZF_{ST} = 10.36; Broiler vs Layer, ZF_{ST} = 7.09) (Fig. S2) is well-known for its role in determining growth rate and body size via regulating the growth hormone (GH) level in blood [76], however, there are no nonsynonymous SNPs in its gene body. IGF2BP1 on chromosome 27 (insulin-like growth factor 2 mRNA-binding protein 1) (Broiler vs RJF, ZF_{ST} = 7.69 and $Z|\Delta\pi|$ = 3.98) (Figs. 3 and 4) may affect growth rate via regulating insulin-like growth factor 2 level [77], but there are no nonsynonymous SNPs in its gene body. The IGF2BP1 locus also overlaps the claw percentage QTLs, shank length QTLs, claw weight QTLs, drumstick and thigh weight QTLs, breastbone crest length QTLs, body weight QTLs, body slope length QTLs and femur bending strength QTLs. The result is consistent with a recent report that mutations in the promoter region of the IGF2BP1 gene can affect chicken body size [18]. In addition, the following genes are also interesting as they overlap white striping QTLs, abdominal fat percentage QTLs, wooden breast QTLs and body weight QTLs, and thus might be related to the large body and fast growth rate of broilers, including OVALX on chromosome 2 (Broiler vs RJF, ZF_{ST} = 9.41 and $Z|\Delta\pi|$ = 5.78; Broiler vs other-breeds, ZF_{ST} = 8.45 and $Z|\Delta\pi|$ = 4.18) (Figs. 3, 4, S2 and S3), RRAS on chromosome 5 (Broiler vs RJF, ZF_{ST} = 14.87 and $Z|\Delta\pi|$ = 7.53; Broiler vs other-breeds, ZF_{ST} = 11.73 and $Z|\Delta\pi|$ = 6.00) (Figs. 3, 4, S2 and S3), SYT8 on chromosome 5 (Broiler vs RJF, ZF_{ST} = 6.90 and $Z|\Delta\pi|$ = 5.87; Broiler vs other-breeds, ZF_{ST} = 6.89 and $Z|\Delta\pi|$ = 3.85) (Figs. 3, 4, S2 and S3), *TNNI*2 on chromosome 5 (Broiler vs RJF, ZF_{ST} = 6.90 and $Z|\Delta\pi|$ = 5.87; Broiler vs other-breeds, ZF_{ST} = 6.89 and $Z|\Delta\pi|$ = 3.85) (Figs. 3, 4, S2 and S3) and FBXO28 on chromosome 3 (Broiler vs RJF, ZF_{ST} = 10.25 and $Z|\Delta\pi|$ = 5.30; Broiler vs other-breeds, ZF_{ST} = 8.38 and $Z|\Delta\pi|$ = 4.90) (Figs. 3, 4, S2 and S3). All these genes either has no nonsynonymous SNPs or the allele frequencies of their nonsynonymous SNPs are very low. Thus, it is highly likely that they might be related to the broilers' traits through changes in their regulatory regions.

For layers, we identified 36 genes in selective windows (Table S39) that might be related to its unique traits including larger number of egg-production. Specifically, gene *NOCT* (Nocturnin), ELF2 (ETS-related transcription factor Elf-2) and *MGARP* (mitochondria localized glutamic acid rich protein) are all located in the same selective sweep on chromosome 4 (Layer vs RJF, ZF_{ST} = 9.84 and $Z|\Delta\pi|$ = 4.61; Layer vs other-breeds, ZF_{ST} = 10.80 and $Z|\Delta\pi|$ = 5.24; Broiler vs Layer, ZF_{ST} = 9.23 and $Z|\Delta\pi|$ = 5.20) (Figs. 3, 4, S2 and S3). *NOCT* is expressed in retina and many other tissues, and its expression shows circadian rhythm [78, 79]. *NOCT* is known to be involved in adipogenesis, osteogenesis, and obesity in mice [80]. It has been shown that *MGARP* is involved in the synthesis of estrogen in ovary, and its expression is under the control of the hypothalamic-pituitary-gonadal (HPG) axis [81]. It has been reported that ELF2 plays a

Wu *et al. BMC Genomics*   (2024) 25:428

Page 15 of 20

role in cell proliferation [82]. Moreover, gene *SLC25A15* on chromosome 1 (mitochondrial ornithine transporter 1) (Layer vs RJF, ZF_{ST} = 7.23 and $Z|\Delta\pi|$ = 3.96; Layer vs other-breeds, ZF_{ST} = 9.68 and $Z|\Delta\pi|$ = 3.82; Broiler vs Layer, ZF_{ST} = 10.52 and $Z|\Delta\pi|$ = 3.21) (Figs. 3, 4, S2 and S3) overlaps the oviduct length QTLs, thus might be related to the egg-production rate of layers. Thus, these four genes might be related to the layers' unique traits. However, the four genes either have no nonsynonymous SNPs or the allele frequencies of their nonsynonymous SNPs are very low. Thus, it is highly likely that they might be related to the high egg-production of layer through changes in their regulatory regions.

Validation of eight genes in DSSs that might be related to body size and egg production rate using RT-qPCR

To test our hypothesis that many genes in DSSs might be related to specific traits through changes in their regulatory sequences in the DSSs, resulting in changes in their expression levels, we used RT-qPCR to quantify the expression levels of eight genes in relevant tissues in chicken breeds with contrast traits.

Specifically, as we inferred that *GHR, GHRHR, IGF2BP1* and *OVALX* might be related to body size, we measured their expression levels in the liver, kidney, leg muscle and breast muscle of Daweishan chickens (small body size) and broilers (large body size). As shown in Fig. 5, consistent with our previous results [76], Daweishan chickens had significantly (p <0.01) higher *GHR* expression levels in the kidney, leg muscle and breast muscle tissues than did the broilers. However, it remains to be elucidated how the higher *GHR* expression level is related to

the small body size of Daweishan chickens. On the other hand, *GHRH*R had significantly higher (p < 0.01) expression levels in the liver, but significantly lower (p <0.05) expression levels in the leg muscle of broilers than in the same tissues of Daweishan chickens (Fig. 5). Moreover, *IGF2BP1* had significantly higher (p < 0.05 or p < 0.01) expression levels in the liver, kidney and breast muscle, but significantly low (p < 0.05) expression levels in the leg muscle of broilers than in the respective same tissues of Daweishan chickens (Fig. 5). Furthermore, *OVALX* had significantly higher (p < 0.05 or p < 0.01) expression levels in the liver and kidney, but significantly lower (p < 0.01) expression levels in the leg muscle of broilers than in the same tissues of Daweishan chickens (Fig. 5). The differential expression levels of *GHRHR, IGF2BP1* and *OVALX* in the liver, kidney and muscles might be related to the different metabolic and growth rates of the two breeds.

In addition, as we predicted that *ELF2, MGARP, NOCT* and *SLC25A15* might be related to the egg production rate, we measured their expression levels in 10 tissues of layers (high-egg production rate) and Wuding chickens (low-egg production rate). As shown in Fig. 6, although *ELF2* had significantly higher (p < 0.05) expression level only in the leg muscle of layers than in the same tissue of Wuding chicken, *MGARP, NOCT* and *SLC25A15* had significantly higher (p < 0.01 or p < 0.05) expression levels in almost all the 10 tissues of layers than in the same tissues of Wuding chickens, except for *MGARP* in the liver, for *NOCT* in pituitary and for *SLC25A15* in breast muscle and retina. Moreover, as the expression of *NOCT* shows circadian rhythm [78, 79], we measured its expression level in 10 tissues of layers and Wuding
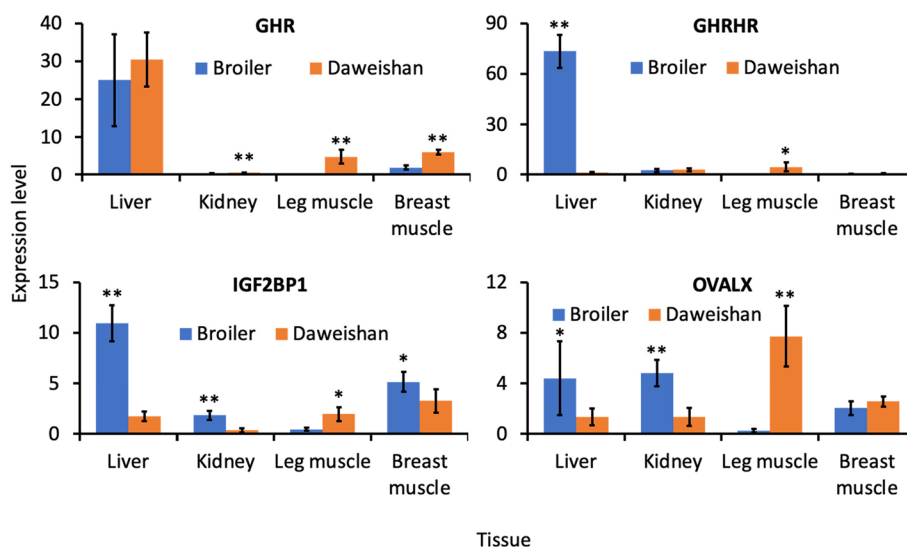

Fig. 5 Comparison of expression levels of *GHR, GHRHR, IGF2BP1* and OVALX in different tissues of broilers and Daweishan chickens. *p < 0.05, **p < 0.01, two-tailed t-test
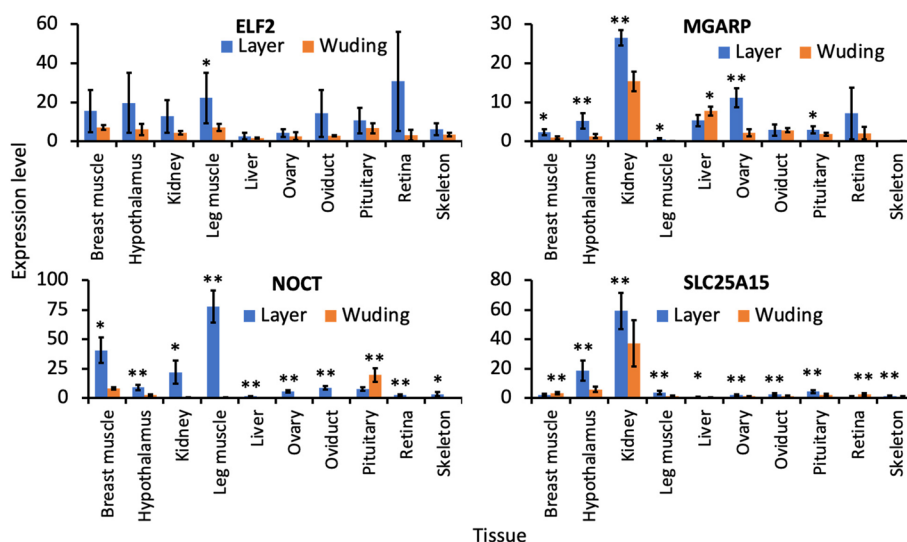
Fig. 6 Comparison of expression levels of *ELF2, MGARP, NOCT* and *SLC25A5* in different tissues of Layers and Wuding chickens. **p* < 0.05, ***p* < 0.01, two-tailed t-test

chickens in 24 hours with 4 hours interval. As shown in Fig. 7, in almost all the 10 tissues, the expression level of *NOCT* was lower after 16:00, and it started to increase at 4:00, peaked at 8:00, and then decreased. In most tissues, layers had significantly higher (p < 0.05 or p < 0.01) expression levels at certain time points than did Wuding chickens. The higher expression level of these four genes in relevant tissues in layers might be related to their high egg-production trait.

Discussion

Next generation sequencing technology makes it possible to re-sequence a large number of individuals for a species for genome-wide studies. In 2021, NCBI released more complete domestic chicken (*Gallus gallus*) genome assemblies (GRCg7b and GRCg7w), providing good reference genomes for this economically, medically and evolutionarily important bird. Using the GRCg7b assembly as the template, we called the variants in populations of five indigenous chicken breeds, broilers, layers and RJFs. By comparing the putative selective sweeps of Daweishan, Hu, Piao, Wuding, Nine-claw chicken, broilers and layers with respect to others and RJFs (19 comparisons, Tables 3 and S11), we identified putative selective sweeps and genes that might be related to the specific traits of each chicken breed or groups (Tables S33 ~ S39). Remarkably, the vast majority (90.5% ~ 98.3%) of our identified DSSs in each of our 19 comparisons overlap QTLs in the chicken QTLdb (Tables 4 and S32), while 74.2% QTLs in the chicken QTLdb overlap our identified DSSs in different comparisons, suggesting that our approach of finding selective sweeps is quite sensitive. Moreover, we also confirm many previously identified genes under
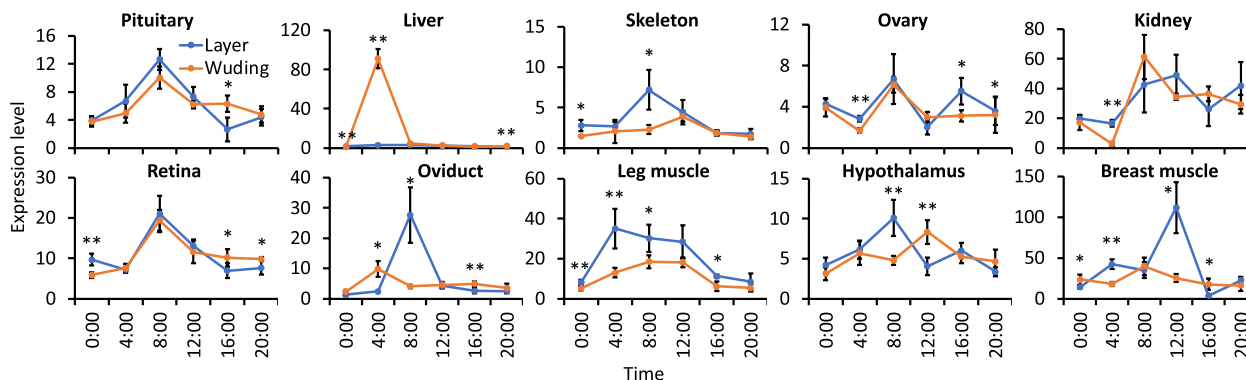

Fig. 7 Comparison of expression levels of *NOCT* at different time of a day in different tissues of layers and Wuding chickens. **p* < 0.05, ***p* < 0.01, two-tailed t-test

artificial selection. Taken together, we have achieved very high prediction precision (or positive prediction values) and sensitivity.

More importantly, our analyses also result in many new findings. Firstly, our results indicate that the RJF A population from northern Thailand are genetically different from those (RJF B) from India. Moreover, the RJF A population are genetically more diverse than the RJF B population due probably to the latter's higher degree of recent inbreeding on the farms since their ancestor's capture as indicted by their higher proportion of long ROH. Both PCA and genetic structure analyses indicate that the five indigenous breeds are closely related to one another, and also more closely related to RJF A than to RJF B. On the other hand, the three layer populations (Layer A, Layer B and Layer C) with different origins have quite different genetic structures though they all are selected for high egg productivity. The same is true for the two broilers populations (Broiler A and Broiler B) though both are selected for high meat productivity. Furthermore, we find that indigenous chickens have higher density of rare allele frequencies for nonsynonymous SNPs than the commercial chickens. As rare alleles tend to be deleterious, the indigenous chickens are more likely to harbor deleterious mutations than commercial chickens. Intensive industrial selective breeding of commercial chickens might lead to the loss of rare alleles which might be slightly deleterious.

Secondly, we identify a much larger number of selective sweeps/DSSs and genes related to the specific traits of broilers and layers than the previous study [52]. We attribute the difference to the different statistic models used in the two studies. More specifically, we use a more rigorous Null model by generating 100-sets of windows with the allele frequencies randomly permutated [6, 51]. Using the mean and standard deviation of the Null model, we compute ZF_{ST} and $Z|\Delta\pi|$ for each window in each comparison. In contrast, the previous study [52] used the mean and standard deviation of the F_{ST} and $|\Delta\pi|$ values of the windows to compute the ZF_{ST} and $Z|\Delta\pi|$, which is not a rigorous Null model. Thus, the previous study might underestimate the number of selective sweeps. Consequently, we identify ~2500 putative DSSs containing ~1800 genes for the broilers and ~2000 putative DSSs containing ~1000 genes for the layers (Tables 3 and S11), which included almost all the only 90 and 66 putative selective sweep windows (40kb) found in broilers and layers, respectively, in the previous study [52]. Notably, we were able to identify more selective sweeps using ZF_{ST} than using $Z|\Delta\pi|$, suggesting that the former might be more sensitive than the latter to identify selective sweeps. In the comparisons between each domestic chicken breed and RJFs, we found numerous genes that might be under selection in both indigenous chickens

and commercial chickens, indicating that these genes might be beneficial for the chickens to live in both traditional and industrial artificial conditions.

Thirdly, we negate several genes reported to be related to the rumpless trait of Piao chicken by a previous study [45] based on our results from multiple comparisons with or without Piao chicken. More specifically, in addition to *IRX*4, the previous study also claimed that *IL18, HSPB2*, and *CRYAB* [45] might be related to the rumpless trait of Piao chicken. As we also find that gene IRX4 is present in putative selective sweeps only in the Piao vs RJF and Piao vs other-breeds comparisons among our 19 comparisons (Tables S14 and S24), it might be related to the rumpless trait of Piao chicken. However, genes IL18, HSPB2, and CRYAB are present in selective sweeps in not only the comparison related to Piao chicken (Piao vs other-breeds, Table S24), but also in comparisons with chickens having a normal tail alone as a group, such as Daweishan, Wuding, Nine-claw chicken, broilers and layers (Tables S12 and S15 ~ S18). Thus, these three genes might not be related to the rumpless trait of Piao chicken.

Fourthly, our analyses provide many novel selective sweeps containing genes that might be related to artificial selection of unique traits of each chicken breed (Tables S33 ~ S39), and some are quite interesting, thus warranting further experimental investigations. For example, it is interesting to experimentally further investigate variations in the regulatory regions of *NOCT* and *MGARP* for their roles in the high egg-production related traits of layers, such as the lack of brooding behaviors, egg-laying circadian rhythm and high demand for light.

Finally, we find that although SNPs in selective sweeps are more likely to alter amino acids than expected, many genes in selective sweeps often lack fixed amino acid-altering mutations. We found that SNPs in non-coding regions in general, or in selective sweeps in particular, are enriched in all the seven domestic chicken breeds as well as RJF analyzed in this study. Thus, most genes in selective sweeps might affect the traits of chicken breeds by changing their expression levels through changes in their cis-regulatory regions. Consistently, we found that all the eight genes (*GHR, GHRHR, IGF2BP1, OVALX, ELF2, MGARP, NOCT, SLC25A15*) that we examined were significantly differentially expressed in relevant tissues of chicken breeds where these genes were in putative selective sweeps from those of chicken breeds where these genes were not in selective sweeps. Particularly, it was recently found that two different forms of deletions in the upstream region of the *IGF2BP1* gene affected body weight of various chicken breeds [18], and we found that the expression levels of the gene in relevant tissues of broilers differed significantly from those in the same tissues of Daweishan chickens (Fig. 5). However, due to the

lack of a more complete map of *cis*-regulatory modules and their constituent transcription factor binding sites in the chicken reference genome, it is difficult to further pin down the sites that affect the expression of these genes and related organism traits. Therefore, it is pressing to map out the *cis*-regulatory elements in the chicken reference genome as has been done for *C. elegans* [83, 84], *D. melanogaster* [83, 85], mice [86, 87] and humans [87, 88].

Conclusions

We identify 30 million single nucleotide variants and small indels in the five indigenous chicken breeds, 10 million of which are novel. Using a rigorous statistic model, we are able to predict substantially more selective sweeps and affected genes than previously reported in both indigenous and commercial breeds. We not only confirm most of previously identified selective sweeps and affected genes in commercial chickens, but also identify numerous novel candidates that might be related to the unique traits of each breed. Most of the variants in selective sweeps are located in non-coding regions and overlap known chicken QTLs, and they might affect the traits of chicken breeds by changing their expression levels through mutations in their *cis*-regulatory elements. Our results can be beneficial to breeding programs in chicken industry for more efficient and secure production of eggs and meat.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10291-5.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Authors' contributions
CG, ZS and JJ supervised and conceived the project; TD, KW, SXY, ZX, YL, ZJ, JZ, RZ, HW, DG, LL, QL and DW collected tissue samples and conducted molecular biology experiments; SW, SSY and ZS performed data analysis; and SW and ZS wrote the manuscript.

Availability of data and materials
All the re-sequencing data of the indigenous chickens are available in the SRA database with the accession number of 'PRJNA893352' (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA893352).

Declarations

Ethics approval and consent to participate
All procedures were conducted in accordance with the Chinese laws on animal experimentation and were approved by the Yunnan Agricultural University Animal Care and Use Committee (approval ID: YAUACUC01), and conducted under the authority of the Project Licence. Animal use and care were in accordance with the ARRIVE guidelines (https://arriveguidelines.org/).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Faculty of Animal Science and Technology, Yunnan Agricultural University, Kunming, Yunnan, China. [2]Department of Bioinformatics and Genomics, The University of North Carolina at Charlotte, Charlotte, NC 28223, USA. [3]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [4]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China.

References
1. Lawal RA, Martin SH, Vanmechelen K, Vereijken A, Silva P, Al-Atiyat RM, et al. The wild species genome ancestry of domestic chickens. BMC Biol. 2020;18(1):13.
2. Fumihito A, Miyake T, Sumi S, Takada M, Ohno S, Kondo N. One subspecies of the red junglefowl (Gallus gallus gallus) suffices as the matriarchic ancestor of all domestic breeds. Proc Natl Acad Sci USA. 1994;91(26):12505–9.
3. Wang MS, Thakur M, Peng MS, Jiang Y, Frantz LAF, Li M, et al. 863 genomes reveal the origin and domestication of chicken. Cell Res. 2020;30(8):693–701.
4. Cheng HW. Breeding of tomorrow's chickens to improve well-being. Poult Sci. 2010;89(4):805–13.
5. Burt DW. Emergence of the chicken as a model organism: implications for agriculture and biology. Poult Sci. 2007;86(7):1460–71.
6. Gheyas AA, Boschiero C, Eory L, Ralph H, Kuo R, Woolliams JA, et al. Functional classification of 15 million SNPs detected from diverse chicken populations. DNA Res. 2015;22(3):205–17.
7. Boschiero C, Moreira GCM, Gheyas AA, Godoy TF, Gasparin G, Mariani P, et al. Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. BMC Genomics. 2018;19(1):83.
8. Zhang H, Wang SZ, Wang ZP, Da Y, Wang N, Hu XX, et al. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC Genomics. 2012;13:704.
9. Guo X, Fang Q, Ma C, Zhou B, Wan Y, Jiang R. Whole-genome resequencing of Xishuangbanna fighting chicken to identify signatures of selection. Genet Sel Evol. 2016;48(1):62.
10. Wang MS, Huo YX, Li Y, Otecko NO, Su LY, Xu HB, et al. Comparative population genomics reveals genetic basis underlying body size of domestic chickens. J Mol Cell Biol. 2016;8(6):542–52.
11. Wang MS, Otecko NO, Wang S, Wu DD, Yang MM, Xu YL, et al. An evolutionary genomic perspective on the breeding of dwarf chickens. Mol Biol Evol. 2017;34(12):3081–8.
12. Zhang M, Yang L, Su Z, Zhu M, Li W, Wu K, et al. Genome-wide scan and analysis of positive selective signatures in dwarf Brown-egg layers and silky fowl chickens. Poult Sci. 2017;96(12):4158–71.
13. Bortoluzzi C, Crooijmans R, Bosse M, Hiemstra SJ, Groenen MAM, Megens HJ. The effects of recent changes in breeding preferences on maintaining traditional Dutch chicken genomic diversity. Heredity (Edinb). 2018;121(6):564–78.

Wu *et al. BMC Genomics*   (2024) 25:428

Page 19 of 20

14. Lawal RA, Al-Atiyat RM, Aljumaah RS, Silva P, Mwacharo JM, Hanotte O. Whole-genome resequencing of red Junglefowl and Indigenous Village chicken reveal new insights on the genome dynamics of the species. Front Genet. 2018;9:264.

15. Wang Q, Li D, Guo A, Li M, Li L, Zhou J, et al. Whole-genome resequencing of Dulong chicken reveal signatures of selection. Br Poult Sci. 2020;61(6):624–31.

16. Ma Y, Gu L, Yang L, Sun C, Xie S, Fang C, et al. Identifying artificial selection signals in the chicken genome. PLoS One. 2018;13(4):e0196215.

17. Rostamzadeh Mahdabi E, Esmailizadeh A, Ayatollahi Mehrgardi A, Asadi Fozi M. A genome-wide scan to identify signatures of selection in two Iranian indigenous chicken ecotypes. Genet Sel Evol. 2021;53(1):72.

18. Wang K, Hu H, Tian Y, Li J, Scheben A, Zhang C, et al. The chicken Pangenome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. Mol Biol Evol. 2021;38(11):5066–81.

19. Kerje S, Carlborg O, Jacobsson L, Schütz K, Hartmann C, Jensen P, et al. The twofold difference in adult size between the red junglefowl and white Leghorn chickens is largely explained by a limited number of QTLs. Anim Genet. 2003;34(4):264–74.

20. McElroy JP, Kim JJ, Harry DE, Brown SR, Dekkers JC, Lamont SJ. Identification of trait loci affecting white meat percentage and other growth and carcass traits in commercial broiler chickens. Poult Sci. 2006;85(4):593–605.

21. Wang Y, Bu L, Cao X, Qu H, Zhang C, Ren J, et al. Genetic dissection of growth traits in a unique chicken advanced intercross line. Front Genet. 2020;11:894.

22. Wang Y, Cao X, Luo C, Sheng Z, Zhang C, Bian C, et al. Multiple ancestral haplotypes harboring regulatory mutations cumulatively contribute to a QTL affecting chicken growth traits. Commun biol. 2020;3(1):472.

23. Zhang H, Shen LY, Xu ZC, Kramer LM, Yu JQ, Zhang XY, et al. Haplotype-based genome-wide association studies for carcass and growth traits in chicken. Poult Sci. 2020;99(5):2349–61.

24. Jennen DG, Vereijken AL, Bovenhuis H, Crooijmans RM, van der Poel JJ, Groenen MA. Confirmation of quantitative trait loci affecting fatness in chickens. Genet Sel Evol. 2005;37(2):215–28.

25. Zhao X, Nie C, Zhang J, Li X, Zhu T, Guan Z, et al. Identification of candidate genomic regions for chicken egg number traits based on genome-wide association study. BMC Genomics. 2021;22(1):610.

26. Yang S, Shi Z, Ou X, Liu G. Whole-genome resequencing reveals genetic indels of feathered-leg traits in domestic chickens. J Genet. 2019;98(2)

27. Bortoluzzi C, Megens HJ, Bosse M, Derks MFL, Dibbits B, Laport K, et al. Parallel genetic origin of foot feathering in birds. Mol Biol Evol. 2020;37(9):2465–76.

28. Li J, Lee M, Davis BW, Lamichhaney S, Dorshorst BJ, Siegel PB, et al. Mutations upstream of the TBX5 and PITX1 transcription factor genes are associated with feathered legs in the domestic chicken. Mol Biol Evol. 2020;37(9):2477–86.

29. Qiu M, Yang C, Du H, Li Q, Zhang Z, Xiong X, et al. Whole-genome resequencing reveals aberrant autosomal SNPs affect chicken feathering rate. Anim Biotechnol. 2022;33(5):884–96.

30. Zhang X, Wang H, Zhang L, Wang Q, Du X, Ge L, et al. Analysis of a genetic factors contributing to feathering phenotype in chickens. Poult Sci. 2018;97(10):3405–13.

31. Zhao J, Yao J, Li F, Yang Z, Sun Z, Qu L, et al. Identification of candidate genes for chicken early- and late-feathering. Poult Sci. 2016;95(7):1498–503.

32. Chen B, Xi S, El-Senousey HK, Zhou M, Cheng D, Chen K, et al. Deletion in KRT75L4 linked to frizzle feather in Xiushui yellow chickens. Anim Genet. 2022;53(1):101–7.

33. Guo X, Li YQ, Wang MS, Wang ZB, Zhang Q, Shao Y, et al. A parallel mechanism underlying frizzle in domestic chickens. J Mol Cell Biol. 2018;10(6):589–91.

34. Huang T, Pu Y, Song C, Sheng Z, Hu X. A quantitative trait locus on chromosome 2 was identified that accounts for a substantial proportion of phenotypic variance of the yellow plumage color in chicken. Poult Sci. 2020;99(6):2902–10.

35. Huang X, Otecko NO, Peng M, Weng Z, Li W, Chen J, et al. Genome-wide genetic structure and selection signatures for color in 10 traditional Chinese yellow-feathered chicken breeds. BMC Genomics. 2020;21(1):316.

36. Nie C, Qu L, Li X, Jiang Z, Wang K, Li H, et al. Genomic regions related to white/black tail feather color in dwarf chickens identified using a genome-wide association study. Front Genet. 2021;12:566047.

37. Zhang GW, Liao Y, Zhang WX, Wu Y, Liu A. A new dominant haplotype of MC1R gene in Chinese black plumage chicken. Anim Genet. 2017;48(5):624.

38. Li D, Sun G, Zhang M, Cao Y, Zhang C, Fu Y, et al. Breeding history and candidate genes responsible for black skin of Xichuan black-bone chicken. BMC Genomics. 2020;21(1):511.

39. Mehlhorn J, Caspers S. The effects of domestication on the brain and behavior of the chicken in the light of evolution. Brain Behav Evol. 2020;95(6):287–301.

40. Zou A, Nadeau K, Wang PW, Lee JY, Guttman DS, Sharif S, et al. Accumulation of genetic variants associated with immunity in the selective breeding of broilers. BMC Genet. 2020;21(1):5.

41. Li J, Lee MO, Davis BW, Wu P, Hsieh Li SM, Chuong CM, Andersson L: The crest phenotype in domestic chicken is caused by a 197 bp duplication in the intron of HOXC10 G3 (Bethesda) 2021, 11(2).

42. Li YD, Liu X, Li ZW, Wang WJ, Li YM, Cao ZP, et al. A combination of genome-wide association study and selection signature analysis dissects the genetic architecture underlying bone traits in chickens. Animal. 2021;15(8):100322.

43. Kondoh D, Ono R, Yanagawa M, Goto T. Morphological variations of caudal skeleton between three chicken breeds. J Vet Med Sci. 2022;84(9):1225–9.

44. Wang Y, Guo F, Qu H, Luo C, Wang J, Shu D. Associations between variants of bone morphogenetic protein 7 gene and growth traits in chickens. Br Poult Sci. 2018;59(3):264–9.

45. Wang YM, Khederzadeh S, Li SR, Otecko NO, Irwin DM, Thakur M, et al. Integrating genomic and transcriptomic data to reveal genetic mechanisms underlying Piao chicken Rumpless trait. Genomics Proteomics Bioinformat. 2021;19(5):787–99.

46. Xu J, Lin S, Gao X, Nie Q, Luo Q, Zhang X. Mapping of id locus for dermal shank melanin in a Chinese indigenous chicken breed. J Genet. 2017;96(6):977–83.

47. Freese NH, Lam BA, Staton M, Scott A, Chapman SC. A novel gain-of-function mutation of the proneural IRX1 and IRX2 genes disrupts axis elongation in the Araucana rumpless chicken. PLoS One. 2014;9(11):e112364.

48. Noorai RE, Freese NH, Wright LM, Chapman SC, Clark LA. Genome-wide association mapping and identification of candidate genes for the rumpless and ear-tufted traits of the Araucana chicken. PLoS One. 2012;7(7):e40974.

49. Chu Q, Yan Z, Zhang J, Usman T, Zhang Y, Liu H, et al. Association of SNP rs80659072 in the ZRS with polydactyly in Beijing you chickens. PLoS One. 2017;12(10):e0185953.

50. Zhang Z, Nie C, Jia Y, Jiang R, Xia H, Lv X, et al. Parallel evolution of polydactyly traits in Chinese and European chickens. PLoS One. 2016;11(2):e0149010.

51. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994;138(3):963–71.

52. Qanbari S, Rubin CJ, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation in modern chicken. PLoS Genet. 2019;15(4):e1007989.

53. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S et al: Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010, 464(7288):587–591 510.1038/nature08832. Epub 2010 Mar 08810.

54. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformat. 2009;25(14):1754–60.

55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformat. 2009;25(16):2078–9.

56. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.

57. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformat. 2019;35(10):1786–8.

58. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformat. 2016;32(11):1749–51.

Wu *et al. BMC Genomics*   *(2024) 25:428*

Page 20 of 20

59. Chen ZL, Meng JM, Cao Y, Yin JL, Fang RQ, Fan SB, et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nat Commun. 2019;10(1):3404.
60. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655–64.
61. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
62. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol. 2016;17(1):122.
63. Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics. 1983;105(3):767–79.
64. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformat. 2011;27(15):2156–8.
65. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. Science. 2006;312(5780):1614–20.
66. Hu ZL, Park CA, Reecy JM. Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. 2022;50(D1):D956–d961.
67. Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, et al. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genet. 2008;4(2):e1000010.
68. Elferink MG, Megens HJ, Vereijken A, Hu X, Crooijmans RP, Groenen MA. Signatures of selection in the genomes of commercial and non-commercial chicken breeds. PLoS One. 2012;7(2):e32720.
69. Shimada M, Tritos NA, Lowell BB, Flier JS, Maratos-Flier E. Mice lacking melanin-concentrating hormone are hypophagic and lean. Nature. 1998;396(6712):670–4.
70. Moon SY, Zheng Y. Rho GTPase-activating proteins in cell regulation. Trends Cell Biol. 2003;13(1):13–22.
71. Nusbaum C, Mikkelsen TS, Zody MC, Asakawa S, Taudien S, Garber M, et al. DNA sequence and analysis of human chromosome 8. Nature. 2006;439(7074):331–5.
72. Girard C, Duprat F, Terrenoire C, Tinel N, Fosset M, Romey G, et al. Genomic and functional characteristics of novel human pancreatic 2P domain K(+) channels. Biochem Biophys Res Commun. 2001;282(1):249–56.
73. Littman DR, Thomas Y, Maddon PJ, Chess L, Axel R. The isolation and sequence of the gene encoding T8: a molecule defining functional classes of T lymphocytes. Cell. 1985;40(2):237–46.
74. Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, et al. The status, quality, and expansion of the NIH full-length cDNA project: the mammalian gene collection (MGC). Genome Res. 2004;14(10b):2121–7.
75. Agarwal SK, Cogburn LA, Burnside J. Dysfunctional growth hormone receptor in a strain of sex-linked dwarf chicken: evidence for a mutation in the intracellular domain. J Endocrinol. 1994;142(3):427–34.
76. Jia J, Ahmed I, Liu L, Liu Y, Xu Z, Duan X, et al. Selection for growth rate and body size have altered the expression profiles of somatotropic axis genes in chickens. PLoS One. 2018;13(4):e0195378.
77. Cao J, Mu Q, Huang H. The roles of insulin-like growth factor 2 mRNA-binding protein 2 in Cancer and Cancer stem cells. Stem Cells Int. 2018;2018:4217259.
78. Baggs JE, Green CB. Nocturnin, a deadenylase in Xenopus laevis retina: a mechanism for posttranscriptional control of circadian-related mRNA. Curr Biol. 2003;13(3):189–98.
79. Stubblefield JJ, Terrien J, Green CB. Nocturnin: at the crossroads of clocks and metabolism. Trends Endocrinol Metab. 2012;23(7):326–33.
80. Hughes KL, Abshire ET, Goldstrohm AC. Regulatory roles of vertebrate Nocturnin: insights and remaining mysteries. RNA Biol. 2018;15(10):1255–67.
81. Zhou M, Wang Y, Qi S, Wang J, Zhang S. The expression of a mitochondria-localized glutamic acid-rich protein (MGARP/OSAP) is under the regulation of the HPG axis. Endocrinol. 2011;152(6):2311–20.
82. Chung IH, Liu H, Lin YH, Chi HC, Huang YH, Yang CC, et al. ChIP-on-chip analysis of thyroid hormone-regulated genes and their physiological significance. Oncotarget. 2016;7(16):22448–59.
83. Kudron MM, Victorsen A, Gevirtzman L, Hillier LW, Fisher WW, Vafeados D, et al. The ModERN resource: genome-wide binding profiles for hundreds of Drosophila and Caenorhabditis elegans transcription factors. Genetics. 2018;208(3):937–49.
84. Van Nostrand EL, Kim SK: Integrative analysis of C. elegans modENCODE ChIP-seq data sets to infer gene regulatory interactions. Genome Res. 2013;23(6):941–53.
85. Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, et al. A cis-regulatory map of the Drosophila genome. Nature. 2011;471(7339):527–31.
86. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012;488(7409):116–20.
87. Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583(7818):699–710.
88. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91–100.

Publisher's Note