



# Whole-Genome Resequencing of Red Junglefowl and Indigenous Village Chicken Reveal New Insights on the Genome Dynamics of the Species

Raman A. Lawal<sup>1\*</sup>, Raed M. Al-Atiyat<sup>2,3</sup>, Riyadh S. Aljumaah<sup>3</sup>, Pradeepa Silva<sup>4</sup>, Joram M. Mwacharo<sup>5</sup> and Olivier Hanotte<sup>1,6\*</sup>

<sup>1</sup> Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, <sup>2</sup> Genetics and Biotechnology, Animal Science Department, Agriculture Faculty, Mutah University, Karak, Jordan, <sup>3</sup> Animal Production Department, College of Food and Agriculture Sciences, King Saud University, Riyadh, Saudi Arabia, <sup>4</sup> Department of Animal Sciences, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka, <sup>5</sup> Small Ruminant Genomics, International Centre for Agricultural Research in the Dry Areas, Addis Ababa, Ethiopia, <sup>6</sup> LiveGene – CTLGH, International Livestock Research Institute, Addis Ababa, Ethiopia

## OPEN ACCESS

### Edited by:

Meng-Hua Li,  
Institute of Zoology (CAS), China

### Reviewed by:

Shikai Liu,  
Ocean University of China, China  
Ed Smith,  
Virginia Tech, United States

### \*Correspondence:

Raman A. Lawal  
lawalakinyanju@yahoo.com  
Olivier Hanotte  
olivier.hanotte@nottingham.ac.uk

### Specialty section:

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

**Received:** 06 March 2018

**Accepted:** 29 June 2018

**Published:** 20 July 2018

### Citation:

Lawal RA, Al-Atiyat RM, Aljumaah RS, Silva P, Mwacharo JM and Hanotte O (2018) Whole-Genome Resequencing of Red Junglefowl and Indigenous Village Chicken Reveal New Insights on the Genome Dynamics of the Species. *Front. Genet.* 9:264. doi: 10.3389/fgene.2018.00264

The red junglefowl *Gallus gallus* is the main progenitor of domestic chicken, the commonest livestock species, outnumbering humans by an approximate ratio of six to one. The genetic control for production traits have been well studied in commercial chicken, but the selection pressures underlying unique adaptation and production to local environments remain largely unknown in indigenous village chicken. Likewise, the genome regions under positive selection in the wild red junglefowl remain untapped. Here, using the pool heterozygosity approach, we analyzed indigenous village chicken populations from Ethiopia, Saudi Arabia, and Sri Lanka, alongside six red junglefowl, for signatures of positive selection across the autosomes. Two red junglefowl candidate selected regions were shared with all domestic chicken populations. Four candidates sweep regions, unique to and shared among all indigenous domestic chicken, were detected. Only one region includes annotated genes (*TSHR* and *GTF2A1*). Candidate regions that were unique to each domestic chicken population with functions relating to adaptation to temperature gradient, production, reproduction and immunity were identified. Our results provide new insights on the consequence of the selection pressures that followed domestication on the genome landscape of the domestic village chicken.

**Keywords:** red junglefowl, *Gallus gallus*, indigenous village chicken, chicken domestication, chicken adaptation, environmental adaptation, positive selection, candidate sweep regions

## INTRODUCTION

Since Charles Darwin proposed a single ancestry of chicken from the red junglefowl, its status as either monophyletic or polyphyletic has been debated (Darwin, 1868; Beebe, 1918; Danforth, 1958; Morejohn, 1968; Fumihito et al., 1994). While the red junglefowl is the main ancestor, some studies are now supporting genetic contributions from other junglefowl species (Eriksson et al., 2008; Lawal, 2017).

Evidences are also controversial as to the timing and places where chicken domestication first occurred (Zeuner, 1963; Crawford, 1984; West and Zhou, 1988; Fumihito et al., 1996; Liu et al., 2006; Xiang et al., 2014, 2015; Peters et al., 2015). A study on mitochondrial DNA suggests multiple centers of chicken domestication (Liu et al., 2006) from which chicken dispersed to different parts of the world through humans' influence. They entered North Africa, the Middle East and Sri Lanka from the Indian subcontinent, while maritime introductions, likely originating initially in South-East Asia, occurred along the coast of East Africa as well as Sri Lanka (Silva et al., 2009; Gifford-Gonzalez and Hanotte, 2011; Mwacharo et al., 2011). Following these migration events, natural and artificial selections have shaped the genome landscape of domestic chicken resulting in a wide spectrum of breeds and ecotypes.

Aside the fancy breeds, domestic chickens primarily come under two major categories; commercial and indigenous village chickens (Schmid et al., 2015). In developing countries, the latter play prominent roles in the livelihood of smallholder farmers, being adapted to their local environmental conditions. They are often under the custody of women and children, mainly kept as dual purpose (eggs and meat) birds. Furthermore, indigenous village chicken showing special visual appeal such as comb type, skin and feather colors may have been selected by smallholder farmers, thereby increasing the frequencies of desirable phenotypes (Dana et al., 2010; Desta et al., 2013). Extensive phenotypic variations such as plumage color and other morphological characteristics, behavioral, and production traits, which are present in domestic chicken but absent in the red junglefowl, are the result of domestication, adaptation to various agro-ecosystems and stringent human selection for production and/or aesthetic values (Schütz et al., 2001; Keeling et al., 2004; Tixier-Boichard et al., 2011).

In commercial chicken lines, the genetic factors that control growth, development, reproduction, and production traits have been well studied (Rubin et al., 2010; Fu et al., 2016). Meanwhile, the genetic mechanisms underlying unique adaptations to tropical environmental pressures and productivity remain poorly studied in indigenous chicken. Likewise, in the red junglefowl, little is known about the genetic control of its adaptation and survival in its natural habitat. Here, we investigate, using whole-genome sequence data, footprints of positive selection in the genome of red junglefowl and domesticated indigenous village chicken in order to better understand the evolutionary pressures during the domestication of the species and its adaptation to different production environments.

## MATERIALS AND METHODS

### Sampling and Sequencing

A total of 27 indigenous domestic village chickens were sampled and then grouped into three populations based on the countries of origin. They include, Ethiopian domestic chicken from two districts, Horro ( $n = 6$ , altitude around 2,320 m above sea level (asl)) and Jarso ( $n = 5$ , altitude of around 1,870 m asl), Saudi Arabian domestic chicken from three villages, Al Qurin ( $n = 2$ , altitude around 130 m asl), Goliglah ( $n = 2$ , altitude around

130 m asl) and Al Oyoum ( $n = 1$ , altitude around 110 m asl) in the Eastern Province, and Sri Lankan domestic chicken from Puttalam district ( $n = 11$ , altitude around 60 m asl). Horro is a sub-humid region, with an annual rainfall of 1,685 mm and an average temperature of around 19°C. Jarso is semi-arid with an average annual temperature of 21°C and annual average rainfall of 700 mm (Desta et al., 2013). The Eastern Province of Saudi Arabia has an average annual temperature of 26°C (ranging from 21.2 to 50.8°C) and average annual rainfall of 74 mm. Puttalam district of Sri Lanka has an average annual rainfall of ~1,000 mm and temperature of 27°C.

Collection of blood samples was through the wing vein and genomic DNA was extracted using ammonium acetate precipitation (Bruford et al., 1998) and phenol-chloroform protocols. A minimum of 3 µg at 30 ng/µl DNA concentration was used for whole genome re-sequencing. Samples were sequenced at the Beijing Genomic Institute (BGI) or at Novogene on a HiSeq 2000/2500 Illumina platform. Five hundred (500) bp paired-end insert size libraries with read lengths of between 90–100 bp and genome coverage of between 10X and 30X (Table S1) were generated. Adapter pollutions from the raw reads and sequences with quality scores  $\leq 5$  were deleted at source BGI/Novogene.

For the six red junglefowl, one whole-genome sequence (15X genome coverage) from a captive bird (Koen Vanmechelen private collection)<sup>1</sup> and five whole genome sequences (12X–36X genome coverage) from the Wang et al. (2015) were included in the analyses (Table S1). The five red junglefowl were sampled in Yunnan (altitude ~3,000 m asl) and Hainan (altitude ~1,840 m asl) provinces, China. Yunnan is a subtropical highland or humid tropical zone with an annual rainfall range of between 600 mm and 2,300 mm, and annual temperature range of between 8 to 27°C. For the humid tropical Hainan province, the average annual rainfall is about 2,000 mm and temperature ranges between 16 and 29°C. Fastq files for all samples newly sequenced in this study have been deposited to NCBI with the SRA accession number SRP142580 or accessible through <https://www.ncbi.nlm.nih.gov/sra/SRP142580>.

### Sequence Alignment and Variants Calling

The 33 whole-genome sequences were independently aligned to *Galgal* 4.0, which has reference genome size of 1.07 Gb (Hillier et al., 2004), using Burrows-Wheeler Aligner (BWA) version 0.7.5a (Li and Durbin, 2010). Sorting the alignment files into coordinate order, marking the duplicate reads and indexing the binary alignment map (bam) files were done using Picard tools version 1.105<sup>2</sup>. Using the genome analysis toolkit (GATK) version 3.4.0 (McKenna et al., 2010; DePristo et al., 2011; Auwera et al., 2013), we performed a two-steps protocol for local realignment around insertions and deletions (indels) to clean up artifacts that arose, during the initial mapping steps, following misalignments. Finally, we applied a quality score recalibration step for each base call to remove any errors carried over during the sequencing.

<sup>1</sup><http://www.ccrp.be/>

<sup>2</sup><http://picard.sourceforge.net>

To call variants, we ran “HaplotypeCaller” from GATK for each sample bam file to create a single-sample “gVCF” using the “-emitRefConfidence GVC” option. We then followed the multi-sample aggregation approach which jointly genotyped variants by merging together, records of all genome data from each population. Using the “-selectType SNP” option along with the “SelectVariants” from GATK, we extracted the SNPs from the raw genotype file before filtering the extracted SNPs using “VariantFiltration.” All investigations were restricted to bi-allelic single nucleotide polymorphisms (SNPs) using bcftools version 1.2 (Li et al., 2009), autosomes (chromosomes 1–28) and the full mitochondrial DNA (mtDNA).

The mapping metrics including the percentage of read pairs that properly mapped to the same chromosome, mean depth coverage, total reads mapped, percentage of the genome with bases covered by at least 5, 10 and 20 reads were calculated using samtools version 0.1.19 (Li et al., 2009). Using Ensembl’s “VEP” version 85 (Aken et al., 2016), we predicted the consequences of the variants while the total number of SNPs in each sample/population were identified using VCFtools version 0.1.11 (Danecek et al., 2011). The “VennDiagram” package (Chen and Boutros, 2011) in R was used to plot the unique and shared SNPs between the domestic chicken and red junglefowl.

## Population Structure and Genetic Differentiation

We removed SNPs in linkage disequilibrium to establish the genetic structure of each population and the relationships between samples using PLINK version 1.9<sup>3</sup>. We then assessed the structure of each population unsupervised, using ADMIXTURE version 1.3.0 (Alexander et al., 2009). Using the default (folds = 5) for cross-validation, we ran the analysis for 10 clusters ( $K$ ). For the principal component analysis (PCA), we ran the smartpca program in eigenstrat version 6.0.1 (Price et al., 2006). The proportion of variance explained by each eigenvector was calculated by dividing the corresponding eigenvalue to the sum of all the eigenvalues.

Genome-wide, nucleotide diversity ( $\pi$ ) and genetic differentiation ( $F_{ST}$ ) were calculated within and between population(s), respectively in 20 kb windows with 10 kb slide using VCFtools version 0.1.11 (Danecek et al., 2011). For  $F_{ST}$ , the pairwise values were calculated between each domestic chicken population and the red junglefowl.

## Mitochondrial DNA Analysis

The full mitochondrial consensus sequence was extracted from the whole genome sequence of each of the 33 samples using “consensus” option in bcftools version 1.2 (Li et al., 2009). Multiple sequence alignment was conducted for the 33 mtDNA genomes using ClustalX version 2.1 (Larkin et al., 2007). To identify the best-fit nucleotide substitution model, we ran jModeltest version 2.1.7 (Darriba et al., 2012). The HKY+I+G model (Hasegawa et al., 1985) was selected as the best, based on the Akaike Information Criterion (AIC), and was subsequently used to construct an unrooted maximum likelihood tree using

phym1 3.0 (Guindon and Gascuel, 2003). The tree was then viewed in MEGA 7.0 (Kumar et al., 2016).

To assess the haplogroup (clade) of each mtDNA sequence, we extracted the first 397 bp hypervariable region (HVR) of the D-loop from the full mitochondrial sequences using as reference mtDNA sequences of Komiyama et al. (2003) (NCBI accession number AB098668) and six haplogroups *sensu* Mwacharo et al. (2011) (Table S2). A haplotype data file including all the 40 HVR of D-loop sequences was generated using DnaSP version 5.1 (Librado and Rozas, 2009) from which the median-joining network was constructed using network 5.0.0.1<sup>4</sup>

## Selective Sweep Analysis

To detect putative selection sweeps, we used the pool heterozygosity ( $H_p$ ) method (Rubin et al., 2010). It was performed using a 20 kb window size with a 10-kb sliding step following the equation:

$$H_p = \frac{2 \sum n_{MAJ} \sum n_{MIN}}{(\sum n_{MAJ} + \sum n_{MIN})^2} \quad (1)$$

Where  $\sum n_{MAJ}$  and  $\sum n_{MIN}$  are the sums of major and minor allele frequencies, respectively for all the SNPs in the 20 kb window. The values for the  $H_p$  calculated for each window size were then Z-transformed using the equation:

$$Z(H_p) = \frac{H_p - \bar{X}(H_p)}{\sigma(H_p)} \quad (2)$$

Where  $\bar{X}$  is the mean and,  $\sigma$  is the standard deviation of  $H_p$ .

A genome-wide score of  $Z(H_p) \leq -4.0$  was taken as the threshold after examining the distribution plot of the  $Z(H_p)$  values (Figures S1A–D). The size of each candidate selective sweep region was calculated by adding the number of overlapping adjacent windows above the genome-wide threshold.

Since the accuracy of detecting selective sweeps depend on the number of SNPs in each window and considering the high polymorphisms identified within populations, only windows with at least 50 SNPs were considered. Following this criterion, 52, 103, 56, and 39 windows were excluded from the Ethiopian, Saudi Arabian and Sri Lankan chicken populations and red junglefowl datasets, respectively.

## Haplotype Trees

In order to assess if a single or multiple haplotypes were selected across population, we build-up haplotype trees for common candidate “domesticated” regions and regions shared between all domestic chicken and red junglefowl. Only shared significant window(s) across population were used to define the region. For this purpose, we included the haplotype sequences from all junglefowl species used in Lawal (2017) study. Maximum likelihood trees were rooted with the green junglefowl and built using Phym1 3.0 (Guindon and Gascuel, 2003) after the evolutionary model was predicted using jModeltest 2.1.7

<sup>3</sup><https://www.cog-genomics.org/plink2>

<sup>4</sup><http://www.fluxus-engineering.com/sharenet.htm>

(Darriba et al., 2012). Genome sequences of the non-red junglefowl species and *G. gallus bankiva* are available at DNA Data Bank Japan Sequence Read Archive (accession no. DRA003951) (Ulfah et al., 2016).

## Remapping the *Galgal* 4.0 Sweep Regions to *Galgal* 5.0 Coordinates

Following the release of the new reference genome *Galgal* 5.0 (Warren et al., 2017), we remapped the *Galgal* 4.0 candidate sweep regions to the corresponding *Galgal* 5.0 coordinates using NCBI remapper (February 2017 release). All the remapping options were set to default threshold. Selective sweep regions based on the *Galgal* 4.0 and their corresponding positions in *Galgal* 5.0 are reported at Tables S4–S7, including changes in the annotated genes between the two reference genomes. Only *Galgal* 5.0 position annotated genes at candidate regions are reported and discussed herein.

## Gene Ontology and Pathways Analysis

To establish the biological significance of the genes found in each candidate selection sweep region, we performed gene ontology and pathways analysis using Database for Annotation, Visualization, and Integrated Discovery (DAVID version 6.8)<sup>5</sup> and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (KOBAS version 3.0)<sup>6</sup>. The Fisher Exact  $P < 0.05$  default threshold was used to identify over-represented genes.

## RESULTS

### Sequencing and SNPs Identification

Following filtering for quality checks and adapter pollutions, clean sequence reads for each domestic chicken sample range between 108.8 and 408.9 million base pairs (bp) depending on the extent of genome coverage (10X–30X) (see Table S1). For each domestic chicken, the number of nucleotides with quality score  $>20$  (Q20) ranged from 94 to 96%.

More than 90% of the read pairs in all samples were properly mapped to the same chromosome. Except for the red junglefowl\_koen sample with 94.69% of mapped reads,  $\geq 97\%$  of all the reads were mapped to the reference genome. On average,  $\geq 97\%$  of the bases were covered by at least 5 reads, while  $\geq 89\%$  of the bases had minimum support of 10 reads (Table S1).

The intermediate genomic variants generated for individual birds using the “HaplotypeCaller” from GATK (Auwera et al., 2013) were used to jointly genotype all samples belonging to a population into a single variants file. Excluding the multi-allelic sites, the average number of SNPs in each sample was  $\sim 6$  million ( $\sim 6$  SNPs/kb). The only exception is red junglefowl5 and red junglefowl\_koen samples having  $\geq 7$  million SNPs. Around 60% of the SNPs were heterozygous in each sample except in three domestic chicken (JB1A25B, JB2A04B, and Saudi Arabia1), which showed  $\sim 45\%$  heterozygous SNPs (Table S1). At the

population level, we identified 13.07 ( $\sim 12$  SNPs/kb), 10.23 ( $\sim 9$  SNPs/kb) and 14.46 ( $\sim 13$  SNPs/kb) million SNPs in Ethiopian, Saudi Arabian, and Sri Lankan domestic chickens, respectively, and 15.31 ( $\sim 14$  SNPs/kb) million SNPs in the red junglefowl. It corresponds to a total of 17.0 million SNPs ( $\sim 16$  SNPs/kb) for the domestic chicken populations combined, and 20.81 million SNPs ( $\sim 19$  SNPs/kb) after combining the genome of all the domestic chicken populations and red junglefowl (Table S3; Figure S2).

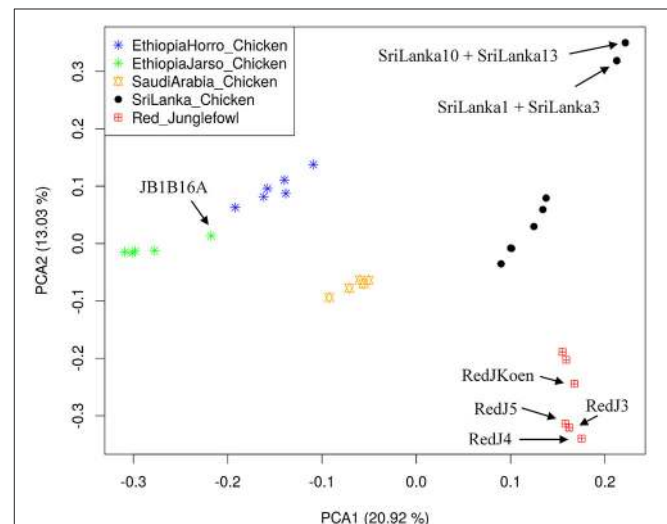
Around 11.05 million SNPs were shared between domestic chicken and red junglefowl, 5.4 and 3.8 million SNPs were unique to domestic chicken and the red junglefowl, respectively (Figure S2). We identified 1.76 million (13% of the total number of SNPs), 1.03 million (10%), and 2.33 million (16%) novel SNPs in Ethiopian, Saudi Arabian and Sri Lankan domestic chickens, respectively and 4.45 million (29%) in the red junglefowl. More than 54% of the SNPs occurred within introns, 30% in intergenic regions, 5.7 and 4.3% in upstream and downstream gene regions, respectively. 3' and 5' UTR variants accounted for 1.8 and 0.4% of the SNPs, respectively (Table S3).

## Population Structure

Population structure at autosomal level was examined using Principal Component (PC) (Figure 1) and Admixture analyses (Figure 2). PC1 and PC2 separate all the domestic populations from the red junglefowl, a result that was also obtained at  $K = 4$  in the admixture analysis. The other admixture plots  $5 \leq K \leq 10$  are shown in Figure S3.

## Diversity and Genetic Differentiation

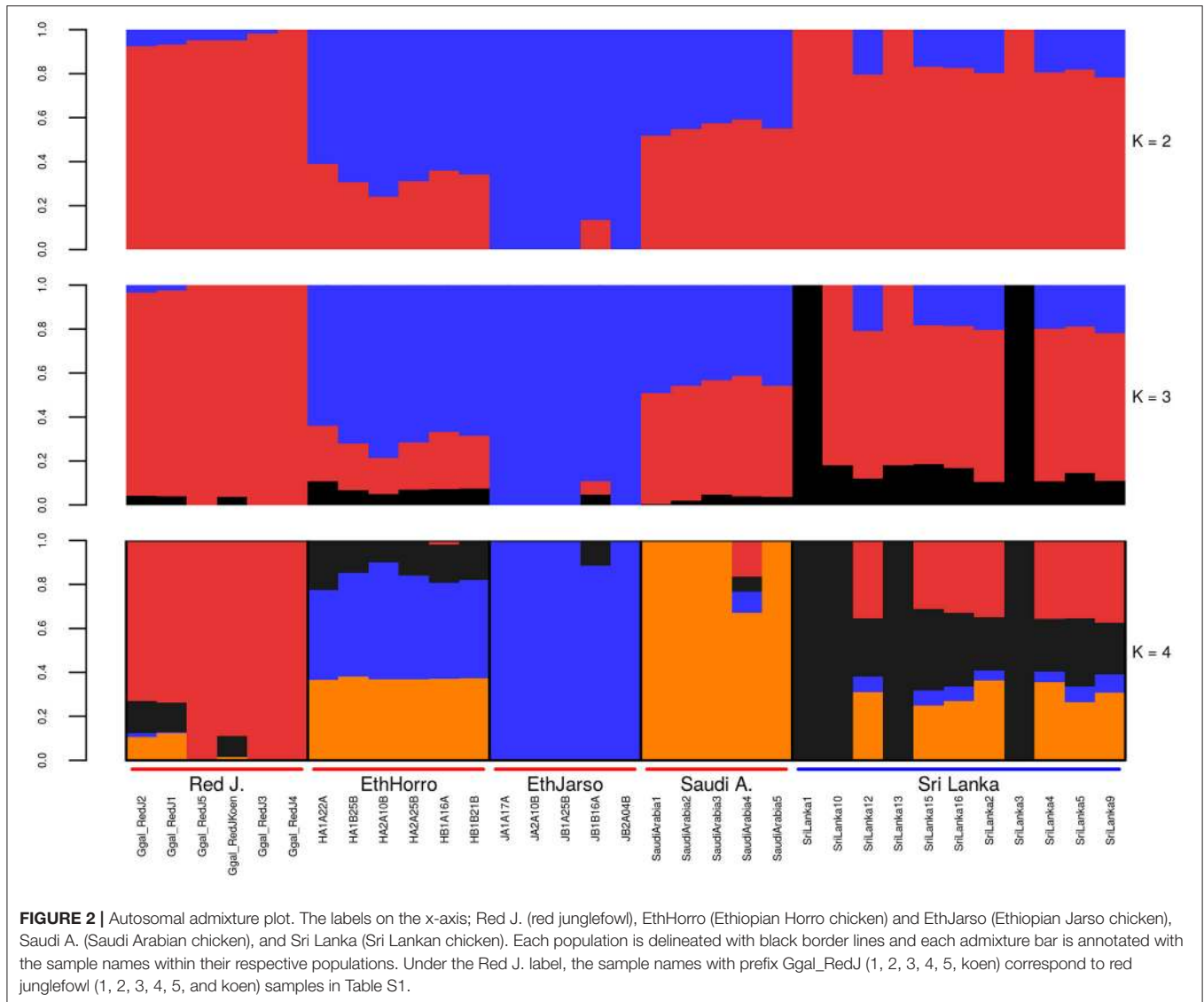
Across populations, we observe the highest genome nucleotide diversity ( $\pi = 0.0052$ ) in the red junglefowl. Among the domestic



**FIGURE 1** | Principal Component Analysis (PCA) plot. The top left label defines colors for each population. Individuals with name annotations have been uniquely identified for comparison purpose with **Figures 2, 3**. The proportion of variance explained by the eigenvector in the x- and y-axes are denoted beside the PCA1 and PCA2.

<sup>5</sup><https://david.ncifcrf.gov/>

<sup>6</sup><http://kobas.cbi.pku.edu.cn/>



chicken populations, Sri Lankan domestic chicken show the highest nucleotide diversity ( $\pi = 0.0046$ ), followed by the Ethiopian Horro ( $\pi = 0.0040$ ), Saudi Arabian ( $\pi = 0.0039$ ) and Ethiopian Jarso domestic chicken ( $\pi = 0.0036$ ).

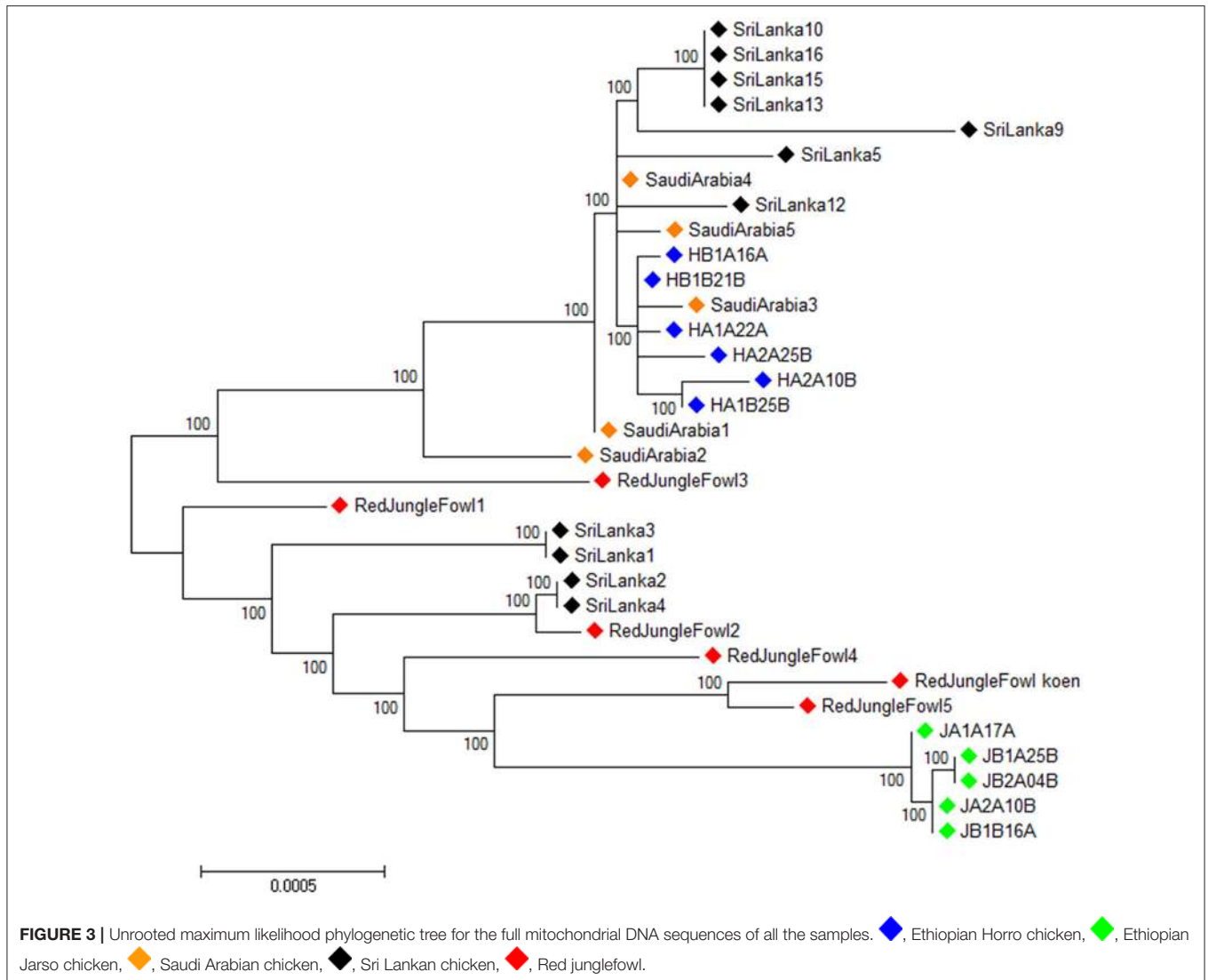
For the pairwise  $F_{ST}$  analysis, we calculated the genetic distances between the red junglefowl and each of the domestic chicken populations to evaluate the levels of autosomal genetic differentiation between domestic chicken and red junglefowl. The Ethiopian Jarso returns the highest  $F_{ST}$  value (0.148), followed by Ethiopian Horro ( $F_{ST} = 0.113$ ), Saudi Arabian ( $F_{ST} = 0.095$ ) and the Sri Lankan domestic chicken ( $F_{ST} = 0.062$ ) populations.

### Mitochondrial Phylogenetic Relationships

The 33 individual mitochondrial genomes were used to construct an unrooted maximum likelihood tree using Phylml 3.0 (Guindon and Gascuel, 2003) (Figure 3). Sri Lankan domestic chicken are divided in two clusters. The first cluster belongs to the same lineage than the Ethiopian Horro and Saudi Arabian chicken. The

second cluster included the red junglefowl and Ethiopian Jarso chicken with the Sri Lankan domestic chicken being closer to the former than the later.

To assess the possible maternal origins of our indigenous village chicken mitochondrial DNA, we extracted the hypervariable region (spanning the first 397 bp) of the mitochondrial DNA D-loop region. We included in our analysis reference haplotypes representing six major chicken haplogroups *sensu* Mwacharo et al. (2011) (Table S2). Haplogroups A, B, C, and D were observed in our dataset (Figure 4). Within a single segregating site, all Ethiopian Horro, four Saudi Arabian and two Sri Lankan domestic chicken haplotypes are linked to haplogroup D. Four Sri Lankan haplotypes are separated by three mutations from the reference D haplotype. Other Sri Lankan domestic chicken haplotypes ( $n = 5$ ) link to haplogroups B and C and a single Saudi haplotype was also close to haplogroup B. The Ethiopian Jarso chicken haplotypes were found closer to haplogroup A.



## Mean Genome Heterozygosity

We calculated the average level of within population  $H_p$  genome heterozygosity (20 kb window size). The genome heterozygosity of the red junglefowl averages to  $0.32 \pm 0.028$  ( $n = 6$ ). Among the domestic chicken populations, Ethiopian chicken population shows the lowest level of genome heterozygosity (mean  $0.31 \pm 0.051$ ,  $n = 11$ ) followed by Sri Lankan chicken population ( $0.32 \pm 0.039$ ,  $n = 11$ ). Saudi Arabian chicken population shows the highest level of genome heterozygosity ( $0.36 \pm 0.048$ ,  $n = 5$ ) (Table 1).

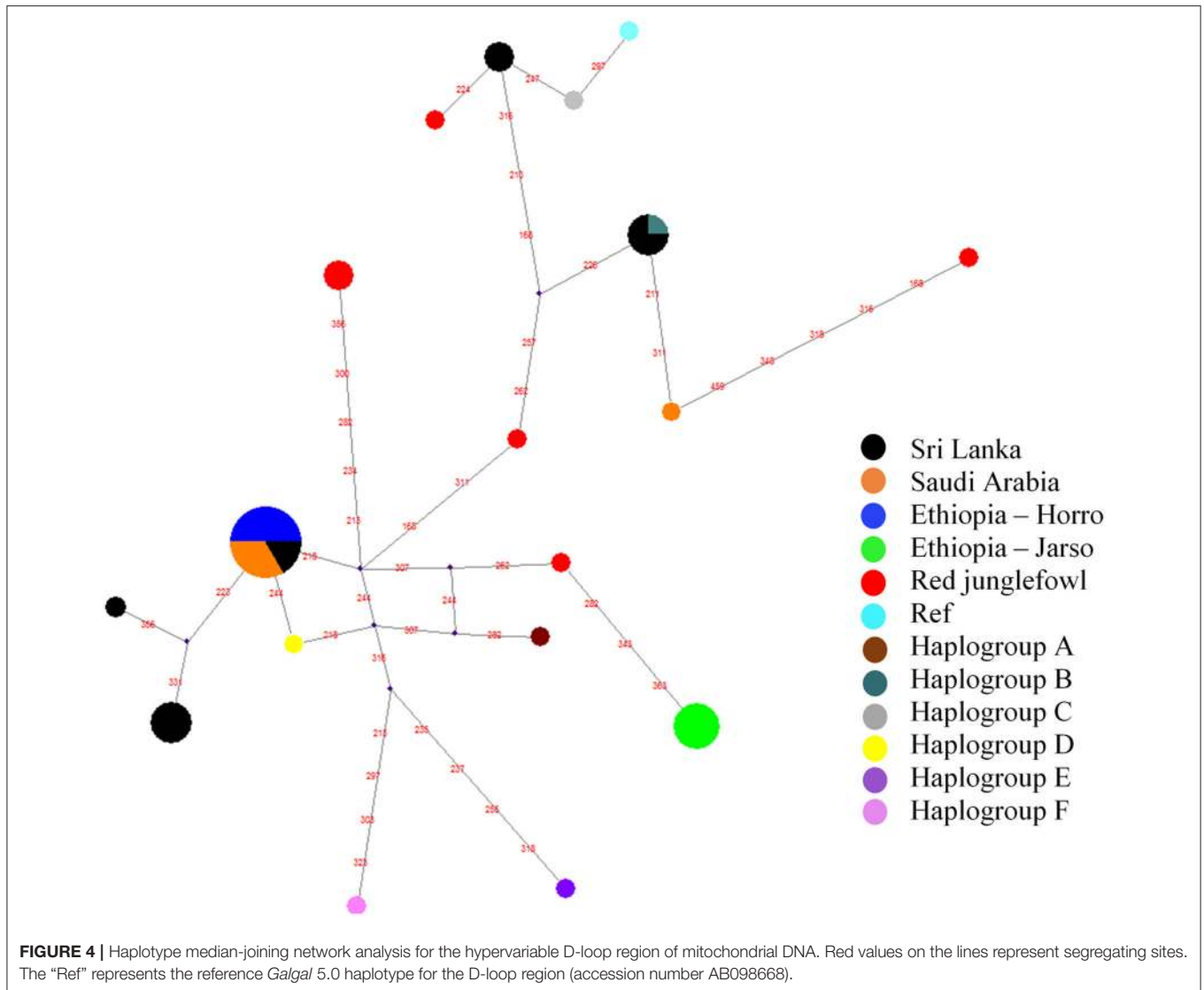
## Selection Sweeps Detection in Red Junglefowl

A total of 434 out of 90,170 windows passed the genome-wide threshold  $\leq -4$  resulting in 190 candidates sweep regions (Table 1; Table S4). Genome-wide, a single  $\sim 20$  kb window located on chromosome 5 (*Galgal* 5.0 position 51895684–51909028 bp) had the lowest  $Z(H_p)$  score ( $-5.93$ ) (Figure 5;

Table S4). The region with the largest fragment size ( $\sim 210$  kb, *Galgal* 5.0 position 2376153–2590429 bp,  $Z(H_p)$  score =  $-4.63 \pm 0.653$ ) is on chromosome 22. Two other candidate regions  $>100$  kb in size are also present;  $\sim 110$  kb region on chromosome 2 (*Galgal* 5.0 position 33529–143341 bp) and  $\sim 150$  kb on chromosome 22 (*Galgal* 5.0 position 578106–728044 bp). Ninety-one candidates sweep regions out of the 190 have fragment sizes of 20 kb, 44 have sizes of 30 kb, 13 have sizes of 40 kb, 17 have sizes of 50 kb, and 25 have sizes of 60 kb and above, respectively. We did not identify any peaks below our threshold on chromosomes 14, 16, 20, 21, 24, 25, 27, and 28 at  $Z(H_p)$  score  $\leq -4$  (Figure 5).

## Selection Sweep Detection in the Domestic Chicken

Out of the 89,443 windows analyzed in Ethiopian domestic chicken, 247 windows passed the genome-wide threshold of  $\leq -4$ . They define 84 candidates sweep regions (Table 1; Table S5). The  $\sim 50$  kb candidate region on chromosome 5 (*Galgal*



5.0 position 40828747–40878736 bp) has the lowest  $Z(H_p)$  score ( $-5.8 \pm 0.289$ ) and spans the *TSHR* and *GTF2A1* genes. Genome-wide, the largest candidate sweep region ( $\sim 210$  kb in size, *Galgal* 5.0 position 424781–634785 bp;  $Z(H_p)$  score =  $-4.29 \pm 0.055$ ) is on chromosome 8 (**Figure 6**; Table S5). Three other candidate regions have fragment sizes  $>100$  kb; two on chromosome 3 with a size of  $\sim 110$  kb (*Galgal* 5.0 position 103157991–103267894 bp) and  $\sim 150$  kb (*Galgal* 5.0 position 103517529–103667817 bp), respectively, and the other on chromosome 8 (*Galgal* 5.0 position 164536–274537 bp) with a size of  $\sim 110$  kb (Table S5). The analysis of the fragment sizes of each sweep region found below the genome-wide threshold of  $Z(H_p) \leq -4$  reveals that 36 candidate regions are 20 kb in size, 13 are 30 kb, nine are 40 kb, ten are 50 kb, and 16 have sizes  $\geq 60$  kb. We did not identify any peaks on chromosomes 6, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, and 28 (**Figure 6**).

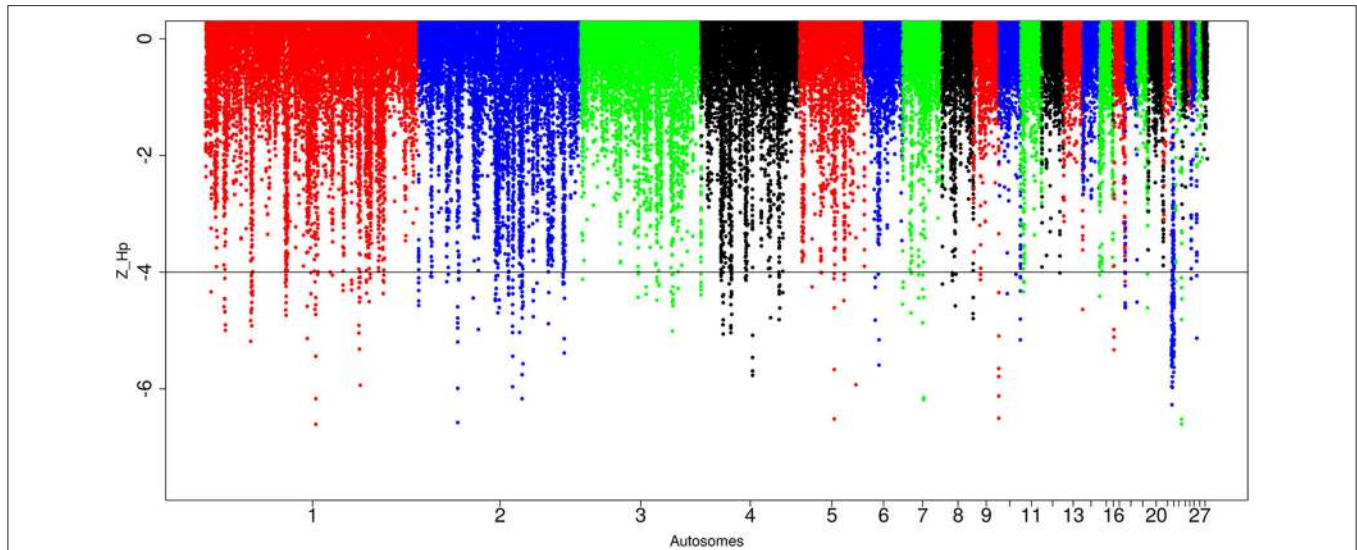
For the Saudi Arabian domestic chicken, we identified in total, 87,646 windows out of which 565 passed the genome-wide

threshold, defining 212 candidate sweep regions (**Table 1**; Table S6). The peak with the lowest  $Z(H_p)$  score ( $-7.27 \pm 0.087$ ) is  $\sim 30$  kb region on chromosome 8 (*Galgal* 5.0 position 204536–234537 bp). The largest sweep region ( $\sim 210$  kb in size, *Galgal* 5.0 position 424781–634785 bp;  $Z(H_p)$  score =  $-4.78 \pm 0.272$ ) occurs on chromosome 8 at the same position as the largest candidate selected region in Ethiopian chicken (**Figure 7**; Table S6). Five other candidate selection sweep regions have sizes  $>100$  kb. It includes two regions on chromosome 2 ( $\sim 140$  kb region at *Galgal* 5.0 position 75375947–75512081 bp, and  $\sim 110$  kb at *Galgal* 5.0 position 147224789–147334917 bp), one region on chromosome 4 ( $\sim 120$  kb in size, *Galgal* 5.0 position 28881313–29001315 bp) and two regions on chromosome 8 ( $\sim 196$  kb length region at *Galgal* 5.0 position 8806310–9002909 bp, and  $\sim 113$  kb region at *Galgal* 5.0 position 9108796–9221862 bp) (Table S6). Analysing fragment sizes for the selection sweep regions show that 81 out of the 212 candidate regions have a fragment size of 20 kb, 55 have a fragment size of 30 kb,

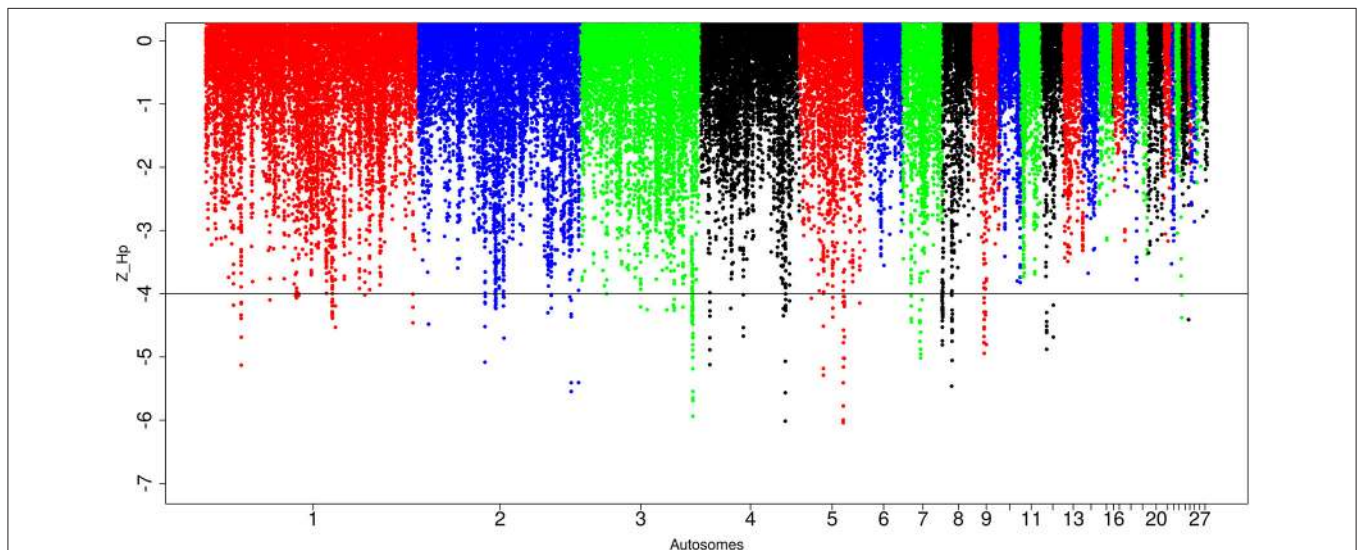
**TABLE 1** | Genome-wide pool heterozygosity ( $H_p$ ) statistics for the three domestic populations and red junglefowl.

Populations	Sample no	Pool heterozygosity ( $H_p$ ) statistics			
		Total number of windows	Genome mean ( $H_p$ )	$Z(H_p) \leq -4.0^a$	Number of candidate sweep regions identified
Ethiopian domestic chicken	11	89,443	$0.31 \pm 0.051$	247	84
Saudi Arabian domestic chicken	5	87,646	$0.36 \pm 0.048$	565	212
Sri Lankan domestic chicken	11	89,701	$0.32 \pm 0.039$	299	127
Red junglefowl	6	90,170	$0.32 \pm 0.028$	434	190

<sup>a</sup>Total number of windows that passed the genome-wide threshold.



**FIGURE 5** | Manhattan plots for selection sweep analysis performed using the standardized pool heterozygosity  $Z(H_p)$  approach. The horizontal line represent the arbitrary threshold for  $Z(H_p) \leq -4$ . This figure shows the selection sweep test for red junglefowl.



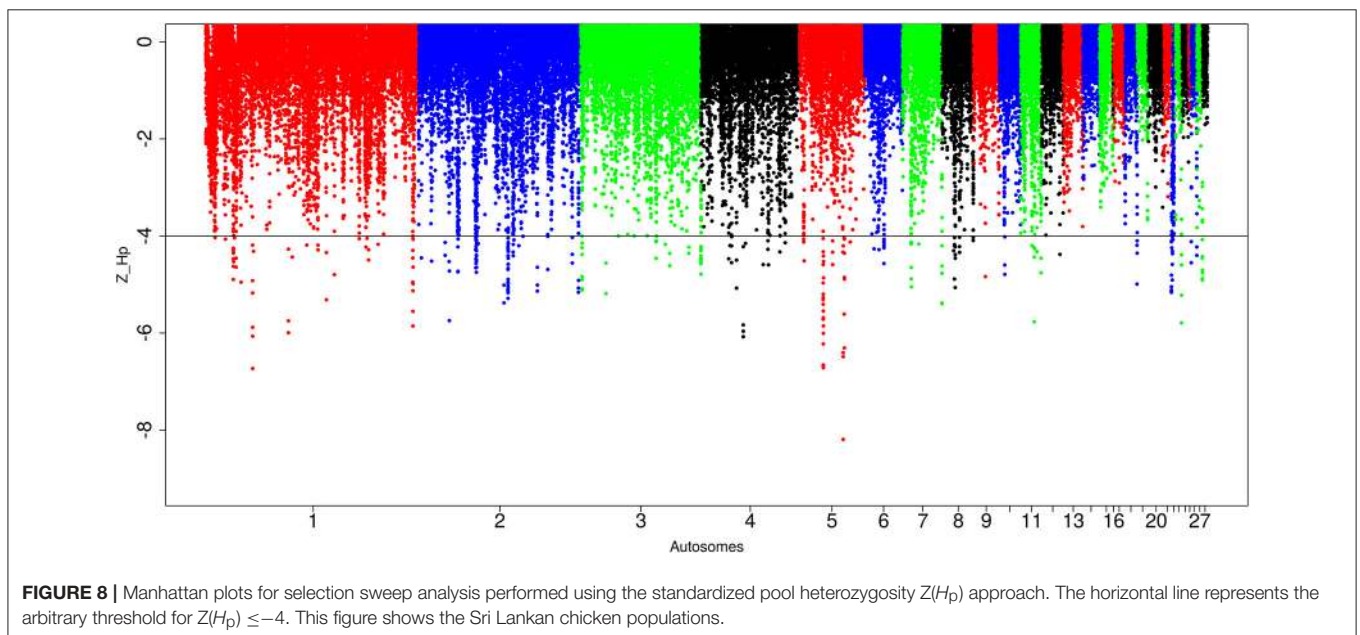
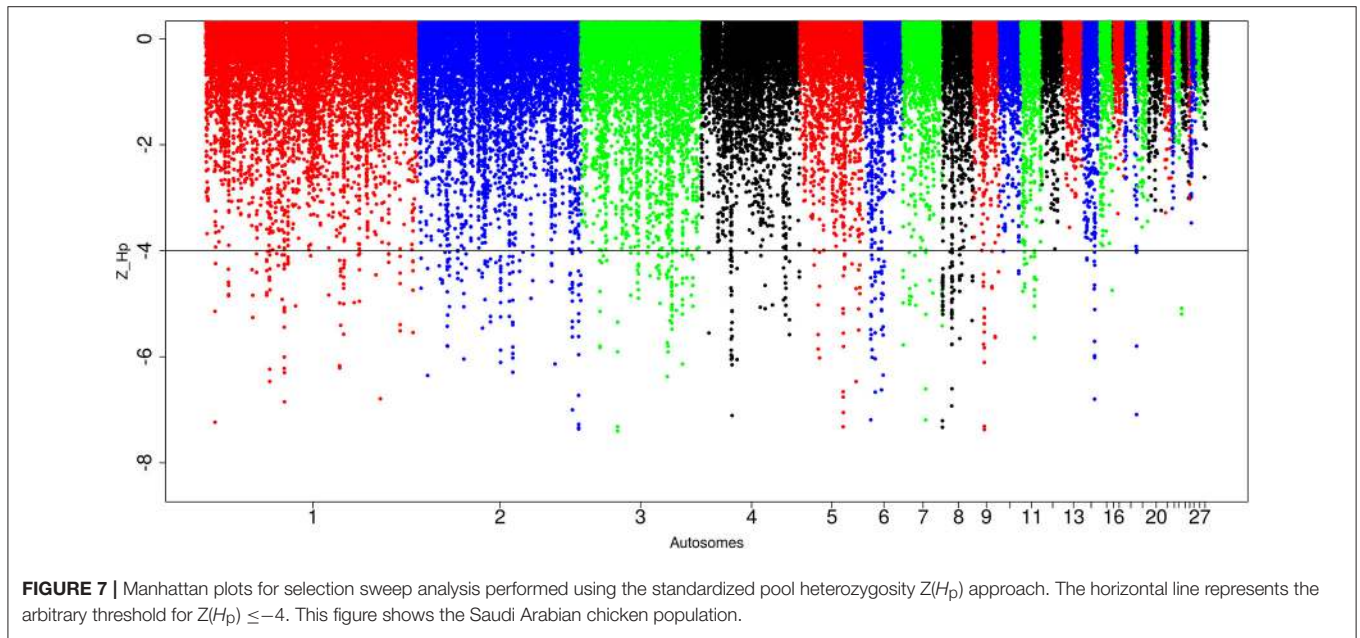
**FIGURE 6** | Manhattan plots for selection sweep analysis performed using the standardized pool heterozygosity  $Z(H_p)$  approach. The horizontal line represents the arbitrary threshold for  $Z(H_p) \leq -4$ . This figure shows the Ethiopian chicken population.



27 are 40 kb in size, 15 are 50 kb size, and 35 are  $\geq 60$  kb in size. We did not identify any peaks below our threshold on chromosomes 12, 13, 16, 17, 19, 20, 21, 22, 24, 25, 26, 27, and 28 (Figure 7).

In Sri Lankan domestic chicken, of the 89,701 windows detected, 299 passed the genome-wide threshold resulting in 127 candidate sweep regions (Table 1; Table S7). Like Ethiopian chicken, the lowest genome-wide  $Z(H_p)$  score ( $-6.32 \pm 1.634$ ) occurs in  $\sim 50$  kb region on chromosome 5 (Galgal 5.0 position 40828747–40878736 bp) (Figure 8; Table S7). The candidate region with the largest fragment size ( $\sim 290$  kb;  $Z(H_p)$

score =  $-4.65 \pm 0.454$ ) is located on chromosome 2 (Galgal 5.0 position 82190953–82481139 bp). Two other candidate regions have fragment sizes  $> 100$  kb. They include a  $\sim 130$  kb region on chromosome 3 (Galgal 5.0 position 111008970–111138863 bp) and a  $\sim 220$  kb region on chromosome 5 (Galgal 5.0 position 22371859–22591888 bp). The analysis of the 127 candidate regions reveal that the length of 63 are 20 kb, 30 are 30 kb, 15 are 40 kb, five are 50 kb, and 14 have sizes  $\geq 60$  kb. We did not identify any peak below our threshold on chromosomes 13, 14, 15, 16, 17, 19, 20, 21, 24, 25, and 28 (Figure 8).



**TABLE 2** | Candidate selection sweep regions shared between/among populations.

Galgal 5.0 reference genome coordinates			Domestic chicken populations			Red junglefowl	Galgal 5.0 reference annotation
Chr	Start	End	Ethiopian	Saudi Arabian	Sri Lankan	Red	Genes
1	8655463	8685324		x	x		-
1	25502468	25522471	x		x		-
1	32471039	32491038	x		x		-
1	58796189	58816190	x	x			-
1	82949734	82969734	x	x			-
1	141722332	141782278		x		x	-
1 <sup>a</sup>	190947207	190967194	x	x	x		-
2	28171924	28211919		x	x		5S_rRNA
2	35402148	35482144		x		x	-
2	60957242	60997244	x	x			-
2	70554878	70585159	x			x	-
2	70821057	70851026	x			x	-
2	71021440	71061441	x	x		x	-
2	78554635	78584572	x		x		-
2	82140679	82160679		x	x		-
2	86485412	86555403		x		x	-
2	86625403	86655403		x		x	-
2	92866622	92886622		x		x	<i>gga-mir-1803</i>
2	96092435	96132454		x	x	x	-
2	139258589	139278588	x	x			-
2	141531206	141601206		x	x		<i>KCNQ3</i>
2	147144279	147164279		x	x		-
2 <sup>a</sup>	147254792	147274793	x	x	x		-
3	53690958	53710958		x		x	-
3	79759035	79779035	x	x			<i>HMG3</i>
3	82484657	82504658		x	x		<i>RIMS1</i>
3	84955044	84985013		x		x	-
3	103517529	103667817	x	x			-
3	111068964	111128966			x	x	-
4	27150530	27174313		x	x		-
4	27853766	27873764	x	x			-
4	28621318	28671317		x	x		-
4	39449745	39489745	x		x		<i>TACR3</i>
4	42031828	42051827		x		x	-
4	76373927	76391801	x	x			<i>LCORL</i>
4	78118133	78168138	x	x			-
4	78407934	78487939	x	x			-
5	22511835	22551887	x		x		-
5 <sup>a</sup>	40828747	40878736	x	x	x		<i>GTF2A1, TSHR</i>
5 <sup>a</sup>	41868268	41908264	x	x	x		-
5	41828256	41848267	x			x	-
5	51895684	51909028		x		x	-
5	55431566	55451566	x	x			<i>C14orf37</i>
6	13734753	13764764			x	x	<i>KCNMA1</i>
6	18316795	18346799		x	x		-
7	492550	522549		x		x	<i>COL5A2</i>
7	8369039	8429044	x		x		-
7 <sup>b</sup>	8578942	8598945	x	x	x	x	-

(Continued)

TABLE 2 | Continued

Galgal 5.0 reference genome coordinates			Domestic chicken populations			Red junglefowl	Galgal 5.0 reference annotation
Chr	Start	End	Ethiopian	Saudi Arabian	Sri Lankan	Red	Genes
7	15343407	15393407			x	x	–
7	36850252	36880269		x	x		BAZ2B
7	36890268	36924725		x	x		NAA20, Mar-07
8	164536	274537	x	x			–
8	424781	634785	x	x			–
8	8824776	8844776	x	x			–
8	8894776	8914776	x	x			–
8	9138797	9221862	x	x			–
8	9511843	9541920	x	x			–
8	17163256	17243252		x	x		–
8	29736834	29766834			x	x	TYW3
9	10389300	10433014	x	x			GK5
9	11499478	11529475	x		x		PLOD2
18	10732550	10752550			x	x	JPT1, SLC16A5, ARMC7
22	238113	318051			x	x	ANTXR1, BMP10, ARHGAP25, GKN2
22	578106	728044			x	x	PPP2R2A, EBF2
22	1068084	1098084			x	x	–
22	1098084	1188031			x	x	STC1
22	1208020	1258019			x	x	LOXL2
22	1536501	1636436			x	x	DUSP26, MAK16, LZTS1, ATP6V1B2, RNF122, TTI2, SLC18A1
23 <sup>b</sup>	5521861	5551860					HPCAL4, TRIT1, MYCL
26	5155940	5205938			x	x	OPTC, PRELP

<sup>a</sup>Significant regions in the three domestic chicken populations.

<sup>b</sup>Significant regions in domestic chicken population and red junglefowl.

<sup>\*</sup>means the candidate region is found selected in the respective population.

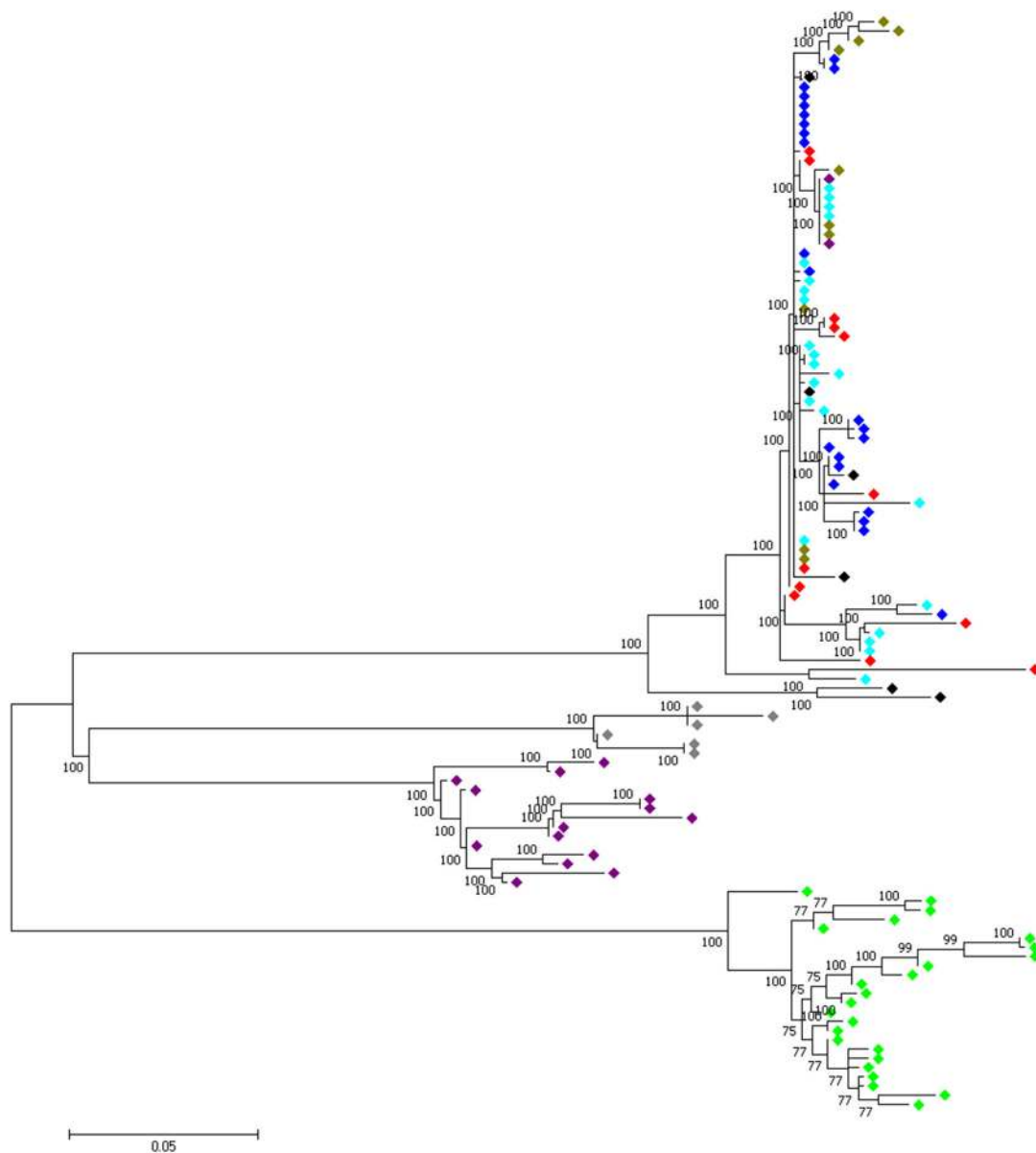
## Overlapping Sweep Regions Across Populations

At the genome level, only two sweep regions are common to all domestic chicken and the red junglefowl. They include ~20 kb candidate region on chromosome 7 (*Galgal* 5.0 position 8578942–8598945 bp) within an intergenic region and ~30 kb length on chromosome 23 (*Galgal* 5.0 position 5521861–5551860 bp) spanning three functional genes (*HPCAL4*, *TRIT1* and *MYCL*) (Table 2). Haplotype trees analysis for the two regions illustrate the variation within the selected haplotypes (Figure 9; Figure S4). One hundred and thirty-two, and 181 variable sites are present across domestic and red junglefowl samples in the 20 and 30 kb regions, respectively (Table 3). It corresponds to an average of 7 and 6 SNPs/kb, well below the combined domestic chicken and red junglefowl populations genome average of 19 SNPs/kb (Figure S2, Table 3).

Four candidate selected regions shared between the three domestic chicken populations are identified. One is located on chromosome 1 (~20 kb: *Galgal* 5.0 position 190947207–190967194 bp), one on chromosome 2 (~20 kb: *Galgal* 5.0 position 147254792–147274793 bp) and two on

chromosome 5 (~50 kb: *Galgal* 5.0 position 40828747–40878736 bp and ~40 kb: *Galgal* 5.0 position 41868268–41908264 bp) (Table 3). We identified two genes (*TSHR* and *GTF2A1*) within the 50 kb region of chromosome 5, while the 20 kb region on chromosome 2 includes an exon of the transcript ENSGALT00000026040. The two other candidate regions are found within intergenic/intronic regions. Figure 10 and Figures S5–S7 illustrates the haplotype variation. Between 179 and 217 variable sites were identified across these regions or an average of 4 to 11 SNPs/kb (Table 3), lower than the genome average of 16 SNPs/kb calculated for the combined domestic chicken populations genomes (Figure S2).

Among the domestic chicken populations, 18 candidates sweep regions, out of a total of 70, are shared between Ethiopian and Saudi Arabian domestic chicken (Table 2). Four of the regions span annotated genes; *HMG3* (chromosome 3), *LCORL* (chromosome 4), *C14orf37* (chromosome 5) and *GK5* (chromosome 9). Two out of the six candidate regions that are shared between the Ethiopian and Sri Lankan domestic chickens overlap with genes including *TACR3* (chromosome 4) and *PLOD5* (chromosome 9). The genes present on the 13



**FIGURE 9** | A 30 kb candidate selected region on chromosome 23 (*Galgal* 5.0 position 5521861–5551860) shared between all domestic chicken population and the red junglefowl. ◆, Ethiopian chicken; ◆, Saudi Arabian chicken; ◆, Sri Lankan chicken; ◆, Red junglefowl; ◆, Javan red junglefowl; ◆, Grey junglefowl; ◆, Ceylon junglefowl; ◆, Green junglefowl.

candidate sweep regions shared between Saudi Arabian and Sri Lankan domestic chickens include *5S\_rRNA* and *KCNQ3* (chromosome 2), *RIMS1* (chromosome 3), *BAZ2B*, *Mar-07* and *NAA20* (chromosome 7) (Table 2).

## Functional Annotations for the Enriched Genes Within the Sweep Regions

To identify the functions of candidate genes that may have played significant roles in adaptation to production environments and the domestication process, we performed enrichment analysis for all genes identified within the candidate sweep regions. Only

classes of genes with default fisher exact  $P < 0.05$  were considered overrepresented for the GO and KEGG pathways analysis. The GO results for all populations is found in Table S8 and that of KEGG pathway is found in Table S9.

## DISCUSSION

The autosomal genetic background and adaptation to local production environments of three populations of indigenous domestic village chicken were analyzed alongside the wild progenitor, the red junglefowl, using whole-genome

**TABLE 3** | Number of variable sites (SNPs) within the selected regions.

	Length (kb)	Total number of SNPs in the selected region	Average SNPs/kb in the selected region*
<b>DOMESTIC CHICKEN SELECTED REGIONS</b>			
1:190947207–190967194	20	217	11
2:147254792–147274793	20	179	9
5:40828747–40878736	50	207	4
5:41868268–41908264	40	212	5
<b>DOMESTIC CHICKEN AND RED JUNGLEFOWL SELECTED REGIONS</b>			
7:8578942–8598945	20	132	7
23:5521861–5551860	30	181	6

\*Average SNPs/kb in the selected region calculated as total number of SNPs in the selected region divided by the length (kb) of the region.

re-sequencing data. Our objectives were to identify candidate positively selected regions (i) shared between wild red junglefowl and domestic chicken, (ii) shared among domestic chicken only and (iii) specific to individual domestic chicken and red junglefowl population.

## Common Genome Regions Selected in Both Domestic and Red Junglefowl

Common regions under selection will be expected between a domesticate and its wild ancestor considering their shared evolutionary history. They may correspond, for examples, to species specific signature of selection underlining shared morphological and behavioral phenotypes. It may be particularly true for village indigenous chicken where human selection pressures have been lower compared to commercial and fancy chicken breeds.

We identified two candidates sweep regions that are shared between all domestic chicken and the red junglefowl. While we could not identify any functional genes within the region on chromosome 7, suggesting possibly an important regulatory role for the region, the one on chromosome 23 spanned three candidate genes (*HPCALA*, *TRITI*, *MYCL*). *HPCALA* is known to play a role in the development of central nervous system (Kobayashi et al., 1998). However, while the biological functions of *MYCL* is still being studied (Brägelmann et al., 2017), both *TRITI* and *MYCL* genes have been linked to the maintenance of tumors (Smaldino et al., 2015; Brägelmann et al., 2017). All three genes may be of importance in both domestic and the wild ancestor; *HPCALA* in relation to behavioral characteristics, *TRITI* and *MYCL* in relation to adaptation to retrovirus infection in particular to virus causing tumors (e.g., leukosis and Marek virus) commonly affecting chicken (Cheng et al., 2010; Wragg et al., 2015).

## Domestic Chicken Specific Signature of Selection

Candidate signature of positive selection specific to domestic chicken may originate from the domestication process itself or after the domestication of the species following geographic dispersion and local responses to human and natural selection

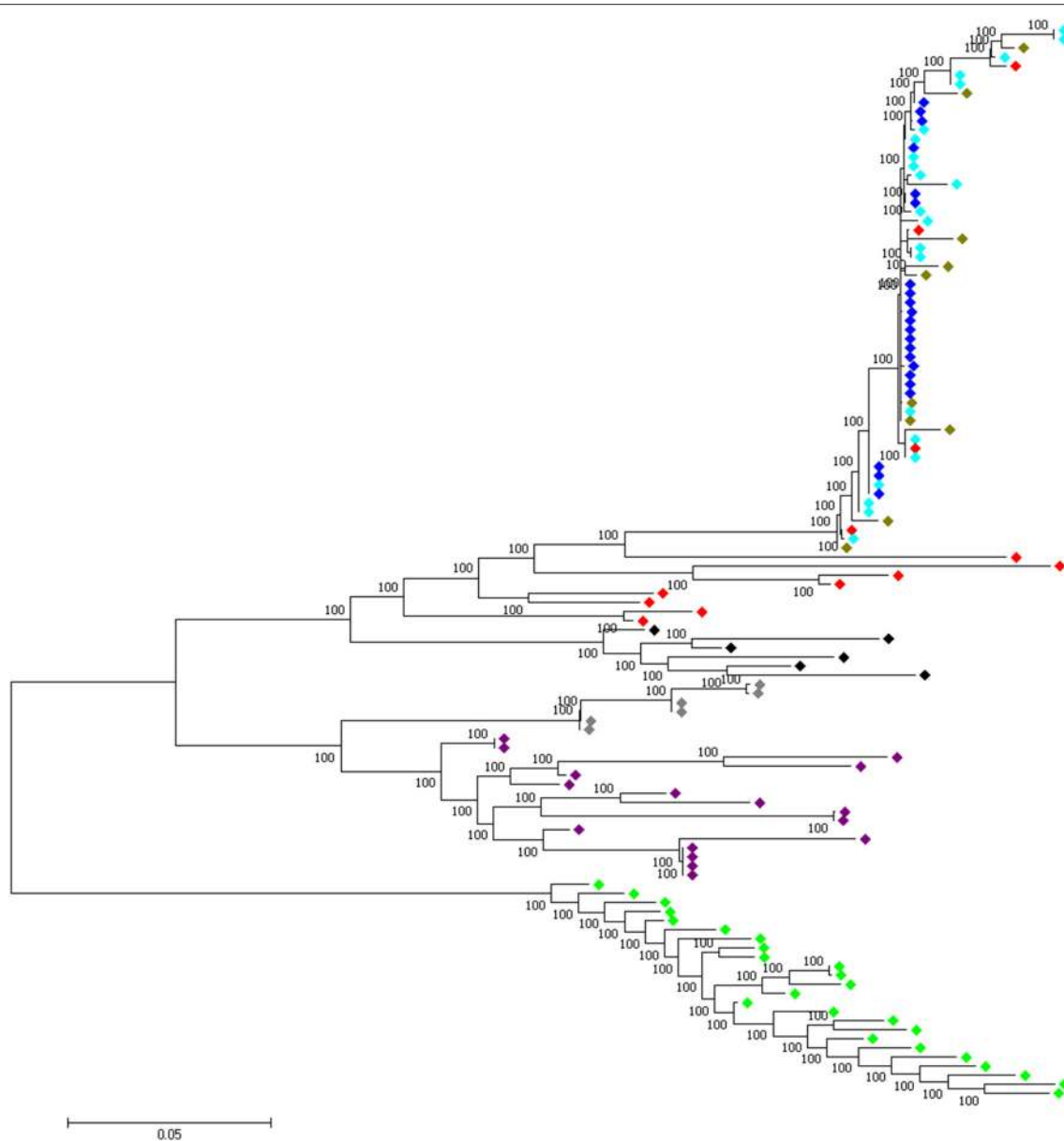
pressures. The distinction between the two is difficult. It may be approached using ancient DNA studies (Flink et al., 2014; Loog et al., 2017). We can also expect that genome regions selected at an early stage of the domestication process, prior to the geographic dispersion of the domesticate will be present in most if not all populations. Compared to fancy chicken breeds and commercial chicken lines, that are characterized by smaller effective population sizes and are heavily selected by humans, the indigenous domestic village chicken, with large effective population sizes, uncontrolled mating and relaxed artificial selection, may represent a better model for the identification of such regions.

We identified four candidate genome regions under positive selection in all the domestic chicken populations but not in the red junglefowl (see **Table 2**). Excluding one region on chromosome 2, these regions have all been previously identified in commercial broilers and layers (Rubin et al., 2010) adding support to early selected domestic region. For the region on chromosome 2, Johnsson et al. (2016) also reported a selected candidate region on this chromosome (*Galgal* 5.0 position 147194251–147234789 bp) which falls 20-kb away from ours (*Galgal* 5.0 position 147254792–147274793 bp). This region is only found in domestic chicken and not in feral birds, and it may be therefore of relevance to the domestication process.

For the remaining three regions, the 50 kb selected region on chromosome 5 includes two genes; the *TSHR* locus involved in metabolic regulation and reproduction process (Yoshimura et al., 2003; Hanon et al., 2008; Rubin et al., 2010) and *GTF2A1*, a candidate biomarker for detecting ovarian tumor (Huang et al., 2009). Hanon et al. (2008) reports that *TSH*-expressing cells of the pars tuberalis is linked to seasonal reproductive control in vertebrates and therefore to the onset of egg laying (Loog et al., 2017). We now know from the studies of Flink et al. (2014) and Loog et al. (2017) that selection at the *TSHR* in European chicken likely followed the selection for higher egg production characteristics. Our studies indicate that similar selection pressures may have acted on Ethiopian, Saudi Arabian, and Sri Lankan domestic chicken. Analysis of chicken populations from different parts of the world, e.g., East and South Asia is required.

## Signatures of Selection in Relation to the Production Environments

Response to selection is environmentally driven either naturally or artificially (Oleksyk et al., 2010). The ancestral species of domestic chicken, the red junglefowl, has a very large geographic range (Delacour, 1977). While different wild red junglefowl subspecies and domestic chicken populations may be witnessing different environmental challenges (e.g., altitudes), all are living in regions that are characterized by rather a warm climate and substantial rainfall which however may show considerable annual variation (e.g., monsoon cycles) or daily variation (e.g., temperature difference during the day). Accordingly, signatures of selection related to thermotolerance including temperature and humidity may be expected in domestic chicken and the red junglefowl.



**FIGURE 10** | A 50 kb candidate selected region on chromosome 5 (*Galgal* 5.0 position 40828747–40878736) identified in all domestic chicken population. This region includes the *TSHR* and *GTF2A1* loci. ◆, Ethiopian chicken; ◆, Saudi Arabian chicken; ◆, Sri Lankan chicken; ◆, Red junglefowl; ◆, Javan red junglefowl; ◆, Grey junglefowl; ◆, Ceylon junglefowl; ◆, Green junglefowl.

In Ethiopian chicken, we identified two candidate genes, *HRH1* and *AGTR1*, associated with “vasoconstriction regulation.” Vasoconstriction has been linked to reduction in peripheral blood flow leading to increase in internal body temperature (Sessler et al., 1990). These genes may likely play important roles in thermoregulation (Collier and Collier, 2011; Su et al., 2011). The reduction in evaporative heat loss and stress through decreased cutaneous blood flow has been reported previously in cattle and birds (Collier and Collier, 2011; Klotz et al., 2016). Compared to the average chicken body temperature of 41°C (Bolzani et al., 1979), the ambient temperatures of

Horro and Jarso districts are relatively low (19 to 21°C) and the two selected candidate genes may played important roles in adaptation to their local environments. At the opposite, Saudi Arabia is very dry with extreme heat during the day which could rise above 50°C in July/August. Here, we identified several GO terms such as “blood circulation,” “regulation of heart contraction,” “regulation of muscle system process,” “regulation of muscle adaptation,” and “regulation of cardiac muscle contraction” that may be linked to the control of blood flow and evaporative cooling (Collier and Collier, 2011). Other studies have associated some of these GO terms to oxygen deprivation

response due to high altitude adaptation (Li et al., 2013; Wang et al., 2015). However, this causative explanation is unlikely in our case because the Saudi Arabian chicken were sampled at an altitude of about 100 m asl. Considering the climatic conditions of the sampling area, we favor here the link to heat loss in response to extreme heat. The significantly enriched GO terms, cellular response to hydrogen peroxide and toll-like receptor signaling pathways, observed in Saudi Arabian chickens may suggest strong selection as well in response to disease challenges (Medzhitov, 2001; Stone and Yang, 2006).

In the genomes of Saudi Arabian and Sri Lankan domestic chicken alongside the red junglefowl, we uncovered the *KCNMA1* gene, that may be linked to hypoxia response challenge. The region harboring this gene did not come as significant in the Ethiopian chicken. *KCNMA1* is associated with the regulation of smooth muscle contraction through the activation of calcium ions (Williams et al., 2004). Increase in calcium ions stimulates hypoxia-inducible factor-1 (Hui et al., 2006). However, the biological roles played by this gene in red junglefowl and Saudi Arabian or Sri Lankan domestic chicken may be different. While in the two domestic chicken populations, it may be related, to heat tolerance and stress control considering the low elevations of the sampling sites; in the red junglefowl however, it may rather play a role in adaptation to high altitudes. Both the Yunnan (altitude ~3,000 m asl) and Hainan (altitude ~1,840 m asl) provinces, where the two red junglefowls were sampled, are mountainous. High elevations are associated with decrease in arterial oxygen content (Simonson et al., 2010). Another gene, *ADAM9*, detected in our red junglefowl, which plays a role in the development of cardiorespiratory system has also been proposed to be involved in adaptation to high-altitude in Tibetan chicken (Zhang et al., 2016).

*KCNMA1* and *ADAM9* were not detected in the candidate regions in Ethiopian chicken. These chickens live at an altitude of around 2,000 m asl. Perhaps, neither the climate and/or altitude where Horro and Jarso populations live result in strong selection pressures in their genomes. Analysis of Ethiopian chicken, living at much higher altitudes may provide further insights on the possible roles of *KCNMA1* and *ADAM9* in altitude adaptation in African domestic chicken.

In addition, one of the previously reported gene under selection in commercial chicken (Rubin et al., 2010; Johnsson et al., 2016), *NT5CIA*, was also identified in the red junglefowl and Sri Lankan indigenous domestic chicken studied here. Importantly, this gene is known to be involved in regulating the levels of heart adenosine during hypoxia and ischemia especially when blood supply becomes inadequate in some parts of the body (Hunsucker et al., 2001). The detection of hypoxia adaptation in both the red junglefowl and domestic chicken may or may not be related to environmental conditions. However, it is well documented that activities relating to extreme exercise may induce hypoxia (Springer et al., 1991; Lindholm and Rundqvist, 2016). Wild and domestic cocks are most often aggressive in nature with the latter having a long history of being selected for cock fighting (Delacour, 1977). We could then argue that the

aggressiveness already presents in the wild relative, due in part to predator evasion and sexual selection behaviors, which can be seen as extreme exercise, may have undergone positive selection in most domestic chicken populations.

## CONCLUSIONS

Examining signature of selection in both domestic chicken and red junglefowl, our study reveals that only two candidate positive selected regions are common to both while four regions are shared across the domestic populations only. Proviso of the relatively low number of red junglefowl examined and the lack of consensus on the geographic origin of the domestic centers of the species, our results illustrate the major impact of human selection activities on the species, and the consequences on the genome landscape of adaptations to new environments. It exemplifies how quickly a domestic species may evolve when under selection pressures in environments.

## AUTHOR CONTRIBUTIONS

RL and OH conceived and designed the project. PS contributed the DNA and provided knowledge on the Sri Lankan chicken. RA-A, RA, and JM provided the Saudi Arabian chicken samples and their genome sequences. RA-A provided knowledge on the Saudi Arabian chicken and the sampling area. RL performed the analyses and OH supervised the project and contributed substantial knowledge on the interpretation of the results. RL prepared and wrote the manuscript. JM and OH revised the manuscript. All the authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This study was conducted during Raman Akinyanju Lawal Ph.D. study which was supported by the University of Nottingham Vice Chancellor's Scholarship (International) award. The Saudi Arabian sampling and genome sequencing was supported through the grant (12-AGR2555-02) from the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia. Sampling of the Ethiopian chicken was supported by a grant jointly sponsored by Biotechnology and Biological Sciences Research Council (BBSRC), the UK Department for International Development (DFID) and the Scottish Government (CIDLID program, BB/H009396/1, BB/H009159/1 and BB/H009051/1). Publication cost was met by CGIAR - Livestock CRP. We also thank Addie Vereijken for providing the Red Junglefowl\_Koen sample.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00264/full#supplementary-material>

## REFERENCES

- Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., et al. (2016). The Ensembl gene annotation system. *Database*. 2016:baw093. doi: 10.1093/database/baw093
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Auwer, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*. 43, 11.10–11.33. doi: 10.1002/0471250953.bi1110s43
- Beebe, W. (1918). *Monograph of the Pheasants*. London, UK: New York Zoological Society. doi: 10.5962/bhl.title.50777
- Bolzani, R., Ruggeri, F., and Olivo, O. (1979). Average normal temperature of the chicken in the morning and after 1-2 days of fasting. *Boll. Soc. Ital. Biol. Sper.* 55, 1618–1622.
- Brägelmann, J., Böhm, S., Guthrie, M. R., Mollaoglu, G., Oliver, T. G., and Sos, M. L. (2017). Family matters: how MYC family oncogenes impact small cell lung cancer. *Cell Cycle* 16, 1489–1498. doi: 10.1080/15384101.2017.1339849
- Bruford, M., Hanotte, O., Brookfield, J., and Burke, T. (1998). *Single-Locus and Multilocus DNA Fingerprinting*. Oxford, UK: IRL Press.
- Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35. doi: 10.1186/1471-2105-12-35
- Cheng, Z., Liu, J., Cui, Z., and Zhang, L. (2010). Tumors Associated with Avian Leukosis Virus Subgroup J in Layer Hens during 2007 to 2009 in China. *J. Vet. Med. Sci.* 72, 1027–1033. doi: 10.1292/jvms.09-0564
- Collier, R. J., and Collier, J. (2011). *Environmental Physiology of Livestock*. Hoboken, NJ: John Wiley & Sons.
- Crawford, R. (1984). “Domestic fowl,” in: *Evolution of Domesticated Animals*, ed L. L. Mason (London: Longman), 298–311.
- Dana, N., Dessie, T., van der Waaij, L. H., and van Arendonk, J. A. (2010). Morphological features of indigenous chicken populations of Ethiopia. *Anim. Genet. Resour.* 46, 11–23. doi: 10.1017/S2078633610000652
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Danforth, C. (1958). *Gallus sonnerati* and the domestic fowl. *J. Hered.* 49, 167–170. doi: 10.1093/oxfordjournals.jhered.a106797
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109
- Darwin, C. (1868). *The Variation of Animals and Plants Under Domestication*. London, UK: John Murray.
- Delacour, J. (1977). *The Pheasant of the World*. Liss; Hants: Saiga Publishing Co. Ltd. Surr GU26 GTD.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* 43, 491–498. doi: 10.1038/ng.806
- Desta, T., Dessie, T., Bettridge, J., Lynch, S., Melese, K., Collins, M., et al. (2013). Signature of artificial selection and ecological landscape on morphological structures of Ethiopian village chickens. *Anim. Genet. Resour.* 52, 17–29. doi: 10.1017/S2078633613000064
- Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., et al. (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.* 4:e1000010. doi: 10.1371/journal.pgen.1000010
- Flink, L. G., Allen, R., Barnett, R., Malmström, H., Peters, J., Eriksson, J., et al. (2014). Establishing the validity of domestication genes using DNA from ancient chickens. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6184–6189. doi: 10.1073/pnas.1308939110
- Fu, W., Lee, W. R., and Abasht, B. (2016). Detection of genomic signatures of recent selection in commercial broiler chickens. *BMC Genet.* 17:122. doi: 10.1186/s12863-016-0430-1
- Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., and Kondo, N. (1994). One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proc. Natl. Acad. Sci. U.S.A.* 91, 12505–12509. doi: 10.1073/pnas.91.26.12505
- Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., et al. (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6792–6795. doi: 10.1073/pnas.93.13.6792
- Gifford-Gonzalez, D., and Hanotte, O. (2011). Domesticating animals in Africa: implications of genetic and archaeological findings. *J. World Prehist.* 24, 1–23. doi: 10.1007/s10963-010-9042-2
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520
- Hanon, E. A., Lincoln, G. A., Fustin, J.-M., Dardente, H., Masson-Pévet, M., Morgan, P. J., et al. (2008). Ancestral TSH mechanism signals summer in a photoperiodic mammal. *Curr. Biol.* 18, 1147–1152. doi: 10.1016/j.cub.2008.06.076
- Hasegawa, M., Kishino, H., and Yano, T. A. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–174. doi: 10.1007/BF02101694
- Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., et al. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716. doi: 10.1038/nature03154
- Huang, Y. W., Jansen, R. A., Fabbri, E., Potter, D., Liyanarachchi, S., Chan, M. W., et al. (2009). Identification of candidate epigenetic biomarkers for ovarian cancer detection. *Oncol. Rep.* 22, 853–861. doi: 10.3892/or\_00000509
- Hui, A. S., Bauer, A. L., Striet, J. B., Schnell, P. O., and Czyzyk-Krzeska, M. F. (2006). Calcium signaling stimulates translation of HIF- $\alpha$  during hypoxia. *FASEB J.* 20, 466–475. doi: 10.1096/fj.05-5086com
- Hunsucker, S. A., Sychala, J., and Mitchell, B. S. (2001). Human Cytosolic 5'-Nucleotidase I characterization and role in nucleoside analog resistance. *J. Biol. Chem.* 276, 10498–10504. doi: 10.1074/jbc.M011218200
- Jonhsson, M., Gering, E., Willis, P., Lopez, S., Van Dorp, L., Hellenthal, G., et al. (2016). Feralisation targets different genomic loci to domestication in the chicken. *Nat. Commun.* 7:12950. doi: 10.1038/ncomms12950
- Keeling, L., Andersson, L., Schütz, K. E., Kerje, S., Fredriksson, R., Carlborg, Ö., et al. (2004). Chicken genomics: feather-pecking and victim pigmentation. *Nature* 431, 645–646. doi: 10.1038/431645a
- Klotz, J. L., Aiken, G. E., Bussard, J. R., Foote, A. P., Harmon, D. L., Goff, B. M., et al. (2016). Vasoactivity and vasoconstriction changes in cattle related to time off toxic endophyte-infected tall fescue. *Toxins* 8:271. doi: 10.3390/toxins8100271
- Kobayashi, M., Sakai, E., Furuta, Y., and Takamatsu, K. (1998). Isolation of two human cDNAs, HLP3 and HLP4, homologous to the neuron-specific calcium-binding protein genes. *DNA Seq.* 9, 171–176. doi: 10.3109/10425179809072192
- Komiyama, T., Ikeo, K., and Gojobori, T. (2003). Where is the origin of the Japanese gamecocks? *Gene* 317, 195–202. doi: 10.1016/S0378-1119(03)00703-0
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Lawal, R. A. (2017). *Signatures of Selection and Introgression in the Genus Gallus*. Ph.D. thesis, University of Nottingham, Nottingham.
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- 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
- Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature Genet.* 45, 1431–1438. doi: 10.1038/ng.2811
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Lindholm, M. E., and Rundqvist, H. (2016). Skeletal muscle hypoxia-inducible factor-1 and exercise. *Exp. Physiol.* 101, 28–32. doi: 10.1113/EP085318



- Liu, Y. P., Wu, G. S., Yao, Y. G., Miao, Y. W., Luikart, G., Baig, M., et al. (2006). Multiple maternal origins of chickens: out of the Asian jungles. *Mol. Phylogenet. Evol.* 38, 12–19. doi: 10.1016/j.ympev.2005.09.014
- Loog, L., Thomas, M. G., Barnett, R., Allen, R., Sykes, N., Paxinos, P. D., et al. (2017). Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Mol. Biol. Evol.* 34, 1981–1990. doi: 10.1093/molbev/msx142
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Medzhitov, R. (2001). Toll-like receptors and innate immunity. *Nat. Rev. Immunol.* 1, 135–145. doi: 10.1038/35100529
- Morejohn, G. V. (1968). Breakdown of isolation mechanisms in two species of captive junglefowl (*Gallus gallus* and *Gallus sonneratii*). *Evolution* 22, 576–582.
- Mwacharo, J. M., Bjørnstad, G., Mobegi, V., Nomura, K., Hanada, H., Amano, T., et al. (2011). Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Mol. Phylogenet. Evol.* 58, 374–382. doi: 10.1016/j.ympev.2010.11.027
- Oleksyk, T. K., Smith, M. W., and O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philos. Trans. Royal Soc. B.* 365, 185–205. doi: 10.1098/rstb.2009.0219
- Peters, J., Lebrasseur, O., Best, J., Miller, H., Fothergill, T., Dobney, K., et al. (2015). Questioning new answers regarding Holocene chicken domestication in China. *Proc. Natl. Acad. Sci. U.S.A.* 12:E2415. doi: 10.1073/pnas.1503579112
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* 38, 904–909. doi: 10.1038/ng1847
- Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R., 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
- Schmid, M., Smith, J., Burt, D. W., Aken, B. L., Antin, P. B., Archibald, A. L., et al. (2015). Third report on chicken genes and chromosomes 2015. *Cytogenet. Genome Res.* 145, 78–179. doi: 10.1159/000430927
- Schütz, K. E., Forkman, B., and Jensen, P. (2001). Domestication effects on foraging strategy, social behaviour and different fear responses: a comparison between the red junglefowl (*Gallus gallus*) and a modern layer strain. *Appl. Anim. Behav. Sci.* 74, 1–14. doi: 10.1016/S0168-1591(01)00156-3
- Sessler, D. I., Moayeri, A., Stoen, R., Glosten, B., Hynson, J., and McGuire, J. (1990). Thermoregulatory vasoconstriction decreases cutaneous heat loss. *Anesthesiology* 73, 656–660. doi: 10.1097/0000542-199010000-00011
- Silva, P., Guan, X., Ho-Shing, O., Jones, J., Xu, J., Hui, D., et al. (2009). Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (*Gallus lafayetti*). *Anim. Genet.* 40, 1–9. doi: 10.1111/j.1365-2052.2008.01783.x
- Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., et al. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science* 329, 72–75. doi: 10.1126/science.1189406
- Smaldino, P. J., Read, D. F., Pratt-Hyatt, M., Hopper, A. K., and Engelke, D. R. (2015). The cytoplasmic and nuclear populations of the eukaryote tRNA-isopentenyl transferase have distinct functions with implications in human cancer. *Gene* 556, 13–18. doi: 10.1016/j.gene.2014.09.049
- Springer, C., Barstow, T. J., Wasserman, K., and Cooper, D. M. (1991). Oxygen uptake and heart rate responses during hypoxic exercise in children and adults. *Med. Sci. Sports Exerc.* 23, 71–79. doi: 10.1249/00005768-199101000-00012
- Stone, J. R., and Yang, S. (2006). Hydrogen peroxide: a signaling messenger. *Antioxid. Redox Signal.* 8, 243–270. doi: 10.1089/ars.2006.8.243
- Su, M. T., Lin, S. H., and Chen, Y. C. (2011). Genetic association studies of angiogenesis-and vasoconstriction-related genes in women with recurrent pregnancy loss: a systematic review and meta-analysis. *Hum. Reprod. Update* 17, 803–812. doi: 10.1093/humupd/dmr027
- Tixier-Boichard, M., Bed'hom, B., and Rognon, X. (2011). Chicken domestication: from archeology to genomics. *C. R. Biol.* 334, 197–204. doi: 10.1016/j.crv.2010.12.012
- Ulfah, M., Kawahara-Miki, R., Farajallah, A., Muladno, M., Dorshorst, B., Martin, A., et al. (2016). Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. *BMC Genomics* 17:320. doi: 10.1186/s12864-016-2652-z
- Wang, M. S., Li, Y., Peng, M. S., Zhong, L., Wang, Z. J., Li, Q. Y., et al. (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol. Biol. Evol.* 32, 1880–1889. doi: 10.1093/molbev/msv071
- Warren, W. C., Hillier, L. W., Tomlinson, C., Minx, P., Kremitzki, M., and Graves, T. (2017). A new chicken genome assembly provides insight into avian genome structure. *G3* 7, 109–117. doi: 10.1534/g3.116.035923
- West, B., and Zhou, B.-X. (1988). Did chickens go north? New evidence for domestication. *J. Archaeol. Sci.* 15, 515–533. doi: 10.1016/0305-4403(88)90080-5
- Williams, S. E., Wootton, P., Mason, H. S., Bould, J., Iles, D. E., Riccardi, D., et al. (2004). Hemoxygenase-2 is an oxygen sensor for a calcium-sensitive potassium channel. *Science* 306, 2093–2097. doi: 10.1126/science.1105010
- Wragg, D., Mason, A. S., Yu, L., Kuo, R., Lawal, R. A., Desta, T. T., et al. (2015). Genome-wide analysis reveals the extent of EAV-HP integration in domestic chicken. *BMC Genomics* 16:784. doi: 10.1186/s12864-015-1954-x
- Xiang, H., Gao, J., Yu, B., Hofreiter, M., and Zhao, X. (2015). Reply to Peters et al.: further discussions confirm early Holocene chicken domestication in northern China. *Proc. Natl. Acad. Sci. U.S.A.* 112, E2416–E2416. doi: 10.1073/pnas.1503956112
- Xiang, H., Gao, J., Yu, B., Zhou, H., Cai, D., Zhang, Y., et al. (2014). Early Holocene chicken domestication in northern China. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17564–17569. doi: 10.1073/pnas.1411882111
- Yoshimura, T., Yasuo, S., Watanabe, M., Iigo, M., Yamamura, T., Hirunagi, K., et al. (2003). Light-induced hormone conversion of T4 to T3 regulates photoperiodic response of gonads in birds. *Nature* 426, 178–181. doi: 10.1038/nature02117
- Zeuner, F. E. (1963). *A History of Domesticated Animals*. London: Hutchinson & Co. (Publishers) Ltd.
- Zhang, Q., Gou, W., Wang, X., Zhang, Y., Ma, J., Zhang, H., et al. (2016). Genome resequencing identifies unique adaptations of tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. *Genome Biol. Evol.* 8, 765–776. doi: 10.1093/gbe/evw032

**Conflict of Interest Statement:** 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 © 2018 Lawal, Al-Atiyat, Aljumaah, Silva, Mwacharo and Hanotte. 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.