Supplementary Materials for

Title: 863 genomes reveal the origin and domestication of chicken

This file includes:

Methods and Materials and Notes
Supplementary information, Figs. S1-S59
Captions for Table S1 (Table S1 as an Excel file)
Supplementary information, Tables S2-S11
Additional References

Supplementary Text

1. Sample collection and genome sequencing
2. Genomic sequence alignment, SNP calling and variant annotation
3. Phylogenetic relationship and population structuring analyses of RJF subspecies
4. Genetic relationship of DCs with RJF subspecies
5. Modeling the origin and migration of DCs using qpGraph and TreeMix programs
6. Likelihood inference of demographic models based on Site Frequency Spectrum (SFS) using fastsimicoal2 program
7. Genome-wide scanning for variants introgressed from other three junglefowl species into DCs
8. Detection of positively selected genes in the domestic chicken
9. Supplementary Figures (1-59)
10. Supplementary Tables (1-11)
11. Additional References

Materials and Methods

1 Sample collection and genome sequencing

In this study, we collected domestic chickens (DCs) and Red Junglefowls (RJFs) with a wide distribution, ranging from West Asia (Iran and Afghanistan), South Asia (India, Pakistan, Bangladesh, and Sri Lanka), Southeast Asia (Thailand, Vietnam and Indonesia) and China (Supplementary information, Fig. S1 and Table S1; Fig. 1a), as well as European breeds (i.e., White Leghorn), covering all postulated geographic regions for chicken domestication. There is no clear genetic maker and/or specific morphological traits that could be used in distinguishing the five RJF subspecies from each other, and also the extent to which these RJF subspecies diverged from each other is also unknown. However, there is a general consensus regarding the present-day distribution range for each RJF subspecies (Supplementary information, Fig. S1). In this context, we collected many samples from multiple locations in the center inhabitation of each RJF subspecies. In our sampling design for RJFs, to potentially avoid samples admixed with domestic chickens, we considered at least three independent sampling locations that were distant from villages for each subspecies. Genomes for 627 DCs and 142 RJFs including *G. g. bankiva* (n=3), *G. g. gallus* (n=6), *G. g. murghi* (n=68), *G. g. jabouillei* (n=23) and *G. g. spadiceus* (n=42), as well as four Ceylon Junglefowls (*G. lafayettei*), two Grey Junglefowls (*G. sonnerati*) and 12 Green Junglefowls (*G. varius*) were sequenced and reported for the first time in this study. Genome sequencing was performed on the Illumina 2500 HiSeq platform except for 68 Indian RJFs (*G. g. murghi*) and 17 Indian chickens, which were sequenced on the NextSeq500 platform at Xcelris Labs Ltd., Ahmedabad, Gujarat in India. More than 5.5 Tb (Tera base pairs) genomic data was generated.

Additionally, 76 genomes reported previously [1-6] were integrated into our study. In total, 863 genomes including 149 RJFs (*G. g. bankiva* (n=3), *G. g. gallus* (n=6), *G. g. murghi* (n=68), *G. g. jabouillei* (n=27) and *G. g. spadiceus* (n=45)), 696 domestic chickens, four Ceylon Junglefowls, two Grey Junglefowls and 12 Green Junglefowls with more than 6.9 Tb genomic data were analyzed in this study (Supplementary information, Table S1). The average nuclear genome sequencing depth for each sample was 6.7×. We also guaranteed that at least one individual of each RJF subspecies was sequenced to an approximate 20× coverage or higher. To examine the mtDNA variations (See Section 4.8; Supplementary information, Fig. S2 and Table S1), our

samples were assigned into different haplotype clusters which cover most of the previously reported haplogroups [7,8], suggesting a high sufficiency of our sampling coverage fit for chicken genetic history inferences.

2 Genomic sequence alignment, SNP calling and variant annotation

We performed filtration procedures to remove low-quality bases/reads using Btrim software [9]. Reads that met the quality filtration and control were mapped onto the chicken reference genome (*Galgal4*) using bwa "BWA-MEM" algorithm (version: 0.7.5; available at https://github.com/lh3/bwa) with default settings except the "-t 8 –M –R <$readgroup>" options [10]. Then, a series of post-processes were carried out to process and filter alignment files, including alignment position sorting, duplicated reads marking and removal, local realignment, and base quality recalibration, using the corresponding tools available in Picard (version 1.56; available at http://picard.sourceforge.net) and Genome Analysis Toolkit (GATK; version 2.6-4) packages [11], including SortSam, MarkDuplicates, RealignerTargetCreator, IndelRealigner and BaseRecalibrator programs. The variants used for was from Ensemble (version 83: ftp://ftp.ensembl.org/pub/release-83/variation/vcf/gallus_gallus/). Lastly, 'depth' function in samtools program (version: 1.3.1; available at http://samtools.sourceforge.net/) was used to calculate genome sequencing depth for each sample.

SNPs were genotyped for all individuals together using UnifiedGenotyper program and further filtered using VariantFiltration command in GATK package ("-T" option). Quality control parameters were set as follows:

- MQ < 25.0
- cluster 3 -window 10
- QUAL < 40.0
- MQ0 ≥ 4 && ((MQ0/(1.0*DP)) > 0.1)

These parameters imply that SNPs with root mean square (RMS) mapping quality less than 25 ("MQ < 25.0"), within clusters (more than 3 SNPs in a 10 bp window; "cluster 3 -window 10") and genotype quality less than 40 ("QUAL < 40.0"), for which reads with zero mapping quality constitute more than 10% of all reads ("MQ0 ≥ 4 && ((MQ0/(1.0*DP)) > 0.1)") at a given site were removed. In addition, we also removed loci with more than two alleles.

In total, ~33.4 million SNPs were identified from our data set, more than those identified in any previous studies [1,2,5,6,12], highlighting the importance and value of the genome resources used in this study. All SNPs were assigned to specific genomic regions and genes using ANNOVAR software [13] according to the ENSEMBL (version: 83) chicken genome annotation information. Most of these variants are falling in intronic and intergenic regions, accounting for ~41.4% and ~53.1%, respectively.

3 Phylogenetic relationship and population structuring analyses of RJF subspecies

3.1 Genealogical tree construction

A total of 149 RJFs comprising all five subspecies (including *G. g. murghi, G. g. jabouillei, G. g. gallus, G. g. spadiceus* and *G. g. bankiva*) sampled in all their natural distribution ranges were incorporated in our study. Currently, there is limited knowledge about their divergence, separation, and evolutionary relationship. First, we constructed an approximately-maximum-likelihood phylogenetic tree using FastTree program (version: 2.1.9; available at http://www.microbesonline.org/fasttree/) [14] with "-fastest" option to explore their phylogenetic relationship. To reduce potential bias arising from putatively missed and/or unreliable genotypes, SNP data set used for phylogenetic reconstruction was filtered using vcftools program [15] with parameters "--maf 0.05 --max-missing 0.9". In addition, SNPs in repeat-masked regions were also removed. Repeat-masked information for chicken reference genome was retrieved from UCSC (http://genome.ucsc.edu/). We constructed trees using genomes of only RJFs, RJFs plus the Green Junglefowls, and RJFs plus the Green, Grey and Ceylon Junglefowls, with *G. g. bankiva* and the Green Junglefowl being used to root the respective trees. The results for these clustering analyses shown in Fig. 1b and Supplementary information, Fig. S3 suggest that RJFs can be primarily classified into five geographically distinct lineages, matching to their relatively separate geographic ranges and corresponding to their taxonomic subspecies classification.

3.2 Principal component analysis (PCA) and population structuring analysis

We performed PCA and genetic structure clustering to measure genetic stratification for the five RJF subspecies. Principal components (PCs) were computed using both GCTA software [16] and

smartpca program from eigensoft package (version: 5.0.2) [17]. Genetic structure clustering was performed using ADMIXTURE program by assuming the number of ancestral populations (*K*) from 2 to 8 [18]. Following the strategy in a previous study [19], ten times with different random seeds were analyzed and these 'Q-matrix' were summarized and compiled with CLUMPAK [20]. The data set we used here was pruned by PLINK toolset (version: 1.90) with parameters "--indep-pairwise 50 10 0.2" [21]

The number of population and the number of individuals for each population might affect the result, we performed PCA and admixture for the following groups:

a. All samples for five RJF subspecies (*G. g. bankiva*, *G. g. gallus*, *G. g. murghi*, *G. g. spadiceus* and *G. g. jaboullei*);
b. All samples for four RJF subspecies (*G. g. gallus*, *G. g. murghi*, *G. g. spadiceus* and *G. g. jaboullei*);
c. Six randomly selected samples for each of *G. g. gallus*, *G. g. murghi*, *G. g. spadiceus* and *G. g. jaboullei*
d. All samples for *G. g. murghi*, *G.g. spadiceus* and *G.g.jaboullei*;
e. 26 randomly selected samples for each of *G. g. murghi*, *G.g. spadiceus* and *G.g.jaboullei*

Consistent with the result from the phylogenetic tree, RJF individuals were clustered into five geographically distinct groups by PC1, PC2 and PC3 according to their classification of subspecies. *G. g. bankiva* shows the highest differentiation from other four RJF subspecies (Supplementary information, Figs. S5-9). *G. g. murghi* has a relatively high diversity, of which a genetic cline is observed across its geographical range, running along northwestern India to northeastern India. Admixture analyses at $K = 2$ separate *G. g. murghi* from other RJF subspecies while all five RJF subspecies begin to have their own structuring patterns when *K* values increase (Supplementary information, Figs. S11-13).

3.3 Differentiation and introgression among RJF subspecies

To measure the genetic differentiation between each pair of RJF subspecies, we computed F_{ST} (fixation index) for each pair of RJF subspecies using smartpca program in eigensoft package (version: 5.0.2) with settings "phylipoutname, fstonly" in par file [17]. To investigate the splitting and admixture of RJF subspecies, the separation and mixture graphs were inferred using

TreeMix (version: 1.13) and qpGraph programs [22-24]. TreeMix was running with "-global" option and qpGraph was ran using default option. Additionally, f_4- and *D*-statistics were also used to measure the introgression among RJF subspecies, which were computed using *fourpop* and qpDstat programs in TreeMix [42] and ADMIXTOOLS packages [22], respectively. The results are shown in Supplementary information, Fig.s S15-16.

3.4 Demographic history of RJF subspecies revealed by high-depth sequenced individual whole genomes

To explore the divergence and splitting of RJF subspecies, we used a multiple sequential Markovian coalescents (MSMC) method [25]. It extends pair-wise sequential Markovian coalescent (PSMC) to estimate population history using more than two haplotypes [26]. Cross-coalescence rate (CCR) was estimated for each pair of RJF subspecies. CCR = 1, > 0 & < 1 and = 0, suggest no split, split with migration and complete split of population pairs, respectively. Genotypes for RJFs were phased together using Beagle software (version: 4.1) [27] and then used as input for running MSMC. To get a more conserved result, we followed the same criteria (CCR = 50%) as used in previous studies [28,29] to define the splitting time for each pair of RJF subspecies. We restricted these analyses to our samples with >20× sequencing coverage. Generation time (g) and mutation rate (μ) for chicken were 1 year and 1.91×10^{-9} substitutions per site per year, respectively. The latter value was estimated using fourfold degenerate sites from bird 1:1 orthologous genes [30].

As shown in Supplementary information, Fig. S17, *G. g. bankiva* is highly divergent and shows an earlier separation from the other four RJF subspecies, at least 0.5 million years ago. Between 50 and 125 kya, *G. g. gallus, G. g. murghi, G. g. jabouillei* and *G. g. spadiceus* slowly began to separate from each other.

4 Genetic relationship of DCs with RJF subspecies

4.1 Phylogenetic trees using individuals as units

Based on whole-genome sequencing data for a large panel of accessions containing 149 RJFs and 696 DCs, we firstly constructed a maximum-likelihood (ML) tree by applying FastTree program [14] to explore which group(s) of RJFs clustered with DCs. Considering the relatively low

average sequencing coverage (around 6.7×) across all genomes, to reduce the potential effect of missed and false genotypes in sequencing and SNP calling procedures on tree construction, highly missed and low frequent variants were first filtered using vcftools program [15] with parameters "--maf 0.05 --max-missing 0.9". And then, SNPs in repeat-masked regions were removed. Repeat-masked information for chicken reference genome was retrieved from UCSC (http://genome.ucsc.edu/). Due to large memory and long computing time needed to run such large number of genomes and associated data set, similar to Der Sarkissian et al.[31], we only selected exon regions for this analysis with 100 bootstrap sampling. The result is shown in Fig. 2a.

4.2 Phylogenetic trees using populations as units

To further infer the patterns of population spliting and admixtures at a population level, we used TreeMix program (version: 1.13) [24] to construct maximum-likelihood tree for all RJF and DC populations. The Green Junglefowl was used as an outgroup. DCs displayed a geographical pattern according to their sampling locations while RJFs grouped according to their classification of subspecies. The phylogenetic tree was firstly built with no migration event (modeled as migration edge) allowed. The data set used for these analyses was filtered using vcftools program with parameters "--maf 0.05 --max-missing 0.9" [15] and further pruned using PLINK toolset (version: 1.90) with parameters "--indep-pairwise 50 10 0.2" [21]. The results are shown in Supplementary information, Fig. S20.

4.3 PCA and population structuring analyses

PCA was used to measure the genetic differentiation among all RJF and DC populations based on the genome-wide SNP data set. It was performed using both GCTA [16] software and smartpca program from eigensoft package (version: 5.0.2) [17]. To limit the potential effects of alleles with low frequency and linkage disequilibrium on the estimation of PCs, the SNP data set was filtered using vcftools program with parameters "–max-missing 0.9" and further pruned by PLINK toolset (version: 1.90) with parameters "--indep-pairwise 50 10 0.2" [21]. Due to the large number of DC and RJF populations included, we did not remove alleles at low frequency. Additionally, ADMIXTURE program [18] was used to infer ancestral component sharing in RJF and DC populations. Following the strategy in a previous study [19] and above, ten times with different

random seeds were analyzed and these 'Q-matrix' were summarized and compiled with CLUMPAK [20]. The number of ancestral components (*K*) was assumed to be from 2 to 18. Both PCA and ADMIXTURE analyses were first performed for all RJFs and DCs, followed by the removal of *G. g. bankiva* and *G. g. gallus* gradually due to their limited numbers of samples and high divergence from other three Red Junglefowl subspecies and all DCs. It is nevertheless expected that such removal could increase the resolution of these analyses. We performed PCA and admixture for following groups:

a. All samples for five RJF subspecies (*G. g. bankiva, G. g. gallus, G. g. murghi, G.g. spadiceus* and *G.g.jaboullei*) and chickens;

b. All samples for four RJF subspecies (*G. g. gallus, G. g. murghi, G.g. spadiceus and G.g.jaboullei*) and chickens;

c. Six randomly selected samples for each of four RJF subspecies (*G. g. gallus, G. g. murghi, G.g. spadiceus and G.g.jaboullei*) and chicken populations (South China, North China, Southwest China, South Asia, Europe, West Asia, and Southeast Asia)

d. All samples for three RJF subspecies (*G. g. murghi, G.g. spadiceus and G.g.jaboullei*) and chickens

e. 10 randomly selected samples for each of for three RJF subspecies (*G. g. murghi, G.g. spadiceus* and *G.g.jaboullei*) and chickens (South China, North China, Southwest China, South Asia, Europe, West Asia, and Southeast Asia)

The results from these analyses are shown in Fig. 2b and Supplementary information, Figs. S21-28. Both GCTA and smartpca generate very similar results and support that all DCs have a closer relationship with *G. g. spadiceus* than with other RJF subspecies. Further, we also pruned variants based on a more strict criterion using PLINK toolset [21], with parameters "--indep-pairwise 50 10 0.1", rerun PCA, and got similar results as that under previous parameters "--indep-pairwise 50 10 0.2". The result of additional PCA (analyzed using SNPs filtered by "--indep-pairwise 50 10 0.1") is not shown here.

4.4 Measuring shared genetic drift using outgroup-f_3 and f_4 statistics

Outgroup-f_3 in form of f_3(outgroup; A, B) measures the number of ancestral alleles shared between two populations (A and B) relative to their common ancestral population (outgroup)

since their divergence [22]. Higher f_3 values suggest more genetic drift shared by A and B, and thus their closer relationship. In order to investigate which RJF subspecies share more genetic drift with DCs, we roughly grouped DCs by their sampling locations and computed outgroup-f_3 statistics in form of f_3(outgroup; Group_RJF, Group_DC). In addition, we also computed outgroup-f_3 in form of f_3 (outgroup; RJF_individual, Group_DC) by separating RJF individuals as single units. Similarly, we also used f_4 statistics in form of f_4(outgroup, DC; RJF1, RJF2) to measure the amount of genetic drift between DCs and two randomly chosen RJF subspecies. Negative f_4-statistics indicate a closer relationship of DC with RJF1 but not with RJF2, and vice versa. Both outgroup-f_3 and f_4 were calculated using *threepop* and *fourpop* functions in TreeMix program (version: 1.13) [24] with the Green Junglefowl being used as an outgroup. Standard errors for these statistics were computed by performing a weighted block jackknife, as default, by *threepop* and *fourpop* functions.

Unsurprisingly, in agreement with the results from trees and PCA, both outgroup-f_3 and f_4 statistics unequivocally demonstrate that *G. g. bankiva* is highly divergent from DCs, while DCs share more genetic drift with *G. g. spadiceus* than other RJF subspecies (Supplementary information, Fig. S29; Fig. 3a). This finding indicates that DCs inherited most of their genetic diversity from *G. g. spadiceus*, reflecting a high probability that DCs originated first from *G. g. spadiceus*.

4.5 Investigating introgression from RJF subspecies into domestic chicken populations

To further investigate the potential introgression of different RJF subspecies into DC populations, we also computed f_4 statistics in the forms of f_4(outgroup, RJF-X; POP_DOMA, POP_DOMB), f_4(outgroup, RJF-X; DOM-X, *G. g. spadiceus*), f_4(outgroup, *DC-X*; *G. g. murghi*, *G. g. gallus*), f_4(outgroup, *DC-X*; *G. g. murghi*, *G. g. jabouillei*) and f_4(outgroup, *DC-X*; *G. g. gallus*, *G. g. jabouillei*). Interestingly, as shown in the results of f_4 and outgroup-f_3 statistics (Supplementary information, Figs. S29, 33-35), all DC populations harbor the greatest ancestry from *G. g. spadiceus* while several DC populations share some extent of genetic drift with particular RJF populations distributed in close proximity to some local DCs. For instance, DCs from South Asia (including India, Pakistan, and Bangladesh) share significantly high genetic components with *G.*

g. murghi; DCs from East Asia carry comparatively high genetic components of *G. g. jabouillei*; and DCs from Indonesia and Thailand own some genetic components of *G. g. gallus*.

To quantitate the level of such localized introgression from RJF subspecies into DCs, we used *qpF4ratio* program in ADMIXTOOLS packages [17]. In this analysis, the admixture proportion in population X derived from B is calculated as: a = f_4(POP_A, POP_O; POP_X, POP_C) / f_4(POP_A, POP_O; POP_B, POP_C); and admixture proportion from A to X is 1-a. Population O represents local DC populations, which were strictly selected from our previously constructed tree, to meet the phylogenetic position required by the program [17]. The results are presented in Supplementary information, Fig. S36.

4.6 Local ancestral inferences

In order to decipher the genetic legacy introgressed from *G. g. murghi* into White Leghorn breed and Indian indigenous chickens, we applied PCAdmix program (version: 1.0) [32], which could infer admixture tracts along individual chromosomes from more than two sources of ancestry. We phased genotypes for all DCs together using Beagle software (version: 4.1) [27], which were further used for running PCAdmix program. In the first calculation, Indian chickens, *G. g. spadiceus* and chickens from Yunnan of China were assumed as possible sources while White Leghorn as a target population. Additionally, we inferred admixture tracts in Indian chickens with *G. g. murghi* when *G. g. spadiceus* and Yunnan chickens were assumed as potential sources. Because there is no genetic recombination map available for chicken, we used a uniform map, which is 2.8 cM/Mb for chr1-9 and 6.4 cM/Mb for chr10-28 and chr32 [33]. The results are presented in Supplementary information, Figs. S37, 47.

4.7 MSMC analysis

We used MSMC to infer the demographic history and population size fluctuation over time for RJFs and DCs [25]. Firstly, we followed the pipeline of samtools program and bamCaller.py script according to the manual to create a genome mask file. Genotypes for all DCs were phased together using Beagle software (version: 4.1) [27]. The mask files and the phased genotypes were then used as input for MSMC analysis. Population size for each group of DCs was estimated using 4 randomly selected haplotypes (2 individuals). We also computed CCR for each pair of

genomes using the same procedures described above. The analyses were repeated a few times with different combinations of population pairs. Generation time (g) and mutation rate (μ) were 1 year and 1.91×10^{-9} substitutions per site per year, respectively [30]. We restricted this analysis using the high-depth sequenced individual genomes.

To further validate the divergent time estimated using MSMC based on CCR, we compared our estimates with published estimates. For example, as reported by Sawai and colleagues [34], DCs diverged from *G. g. gallus* around 58,000±16,000 years ago, which is comparable to our estimation (Fig. 3b; Supplementary information, Fig. S30).

4.8 Mitochondrial DNA haplotype assignment

Due to sequencing depth and coverage varied across samples, it was unable to assemble mtDNA genomes for all samples. We followed the reported strategy [35] to retrieve the mitochondrial DNA (mtDNA) coding region (np 1233-16785 of reference sequence AP003321) variants from the whole-genome sequencing data. We scored the variants to AP003321[36] and then input them into MitoToolPy (mitotoolpy-var.py)[37] to classify the haplogroups based on the diagnostic mutations. The haplogroups for each of the mtDNA sequences of our samples including both RJFs and DCs were under the standardized hierarchical haplogroup nomenclature system of DomeTree[37]. We further manually checked the results of haplo-grouping, especially referring to our recent update of chicken mtDNA haplogroup CDV[7]. Results are shown in Supplementary information, Table S1 and Figure S2.

5 Modeling the origin and migration of DCs using qpGraph and TreeMix programs

5.1 Modeling the history of DCs using qpGraph program

To assess the fit of separation and admixture graph models of DCs, we used qpGraph program [22,23]. It measures the allele frequency correlation patterns of f_2, f_3 and f_4 statistics in tested populations. This is a robust method wildly used to infer the separation and admixture of different human populations [23,38]. We tested the single and multiple origins of DCs from *G. g. spadiceus* and/or *G. g. murghi*, *G. g. gallus* and *G. g. jabouillei* separately. We focused on a set of populations representing broad genetic and geographic diversity, as follows:

- Outgroup: The Green Junglefowl

- RJFs: *G. g. spadiceus*, *G. g. murghi*, *G. g. gallus* and *G. g. jabouillei*
- Chicken in the basal of a phylogenic tree: Yunnan chickens from China
- South Asia: Indian chickens
- Southeast Asia: Thai chickens from Thailand, Vietnamese chickens
- South China: Silkie chickens

Among these RJF and DC populations, *G. g. jabouillei* has an overlapping distribution range with Silkie and Vietnamese chickens; *G. g. murghi* with Indian chickens; and *G. g. gallus* with Southeast Asian chickens; and *G. g. spadiceus* with Yunnan chickens. To simplify the model tested in this analysis, in each run, we started with a skeleton phylogenic tree consisting of the Green junglefowl, *G. g. spadiceus*, Thai and Yunnan chickens, feeding other RJF subspecies and local chicken pairs gradually, and then tested the possibility of single or multiple origin scenarios. The results are displayed in Supplementary information, Figs. S38, 42, 43.

5.2 Inferring admixture graph using TreeMix program

We also used an alternative inference method in TreeMix program [24], to model the splitting and admixture graphs for DCs and RJF subspecies. It could mitigate the confounding effects of admixture with a maximum-likelihood tree. Since all DC and RJF populations consisted of different numbers of individuals, especially for *G. g. gallus* (n=6) and *G. g. bankiva* (n=3) having limited number sizes, the sample size correction was allowed. As Indian chickens are highly admixed with *G. g. murghi* and carries around 18% of *G. g. murghi* ancestry, we included such a priori specified event. A round of optimization was performed based on the original migration edge (option –climb and -cor_mig). Standard errors were estimated in blocks with 1000 SNPs in each independent run (option –k 1000). Indigenous chickens were grouped into three main groups: South Asia (Bangladesh and India), China (Peking, Shandong, Liaoning, Shanxi, Jiangsu and Jiangxi provinces) and Southeast Asia and neighboring region (Yunnan and Guangxi provinces of China, Thailand and Vietnam). TreeMix was run by assuming 0 to 7 migration edges. To visualize trees and residuals from each model, we used a R script *plotting_funcs.R* from TreeMix program [24]. The results are shown in Supplementary information, Fig. S41.

6 Likelihood inference of demographic models based on Site Frequency Spectrum (SFS) using fastsimicoal2 program

Comparing the observed joint SFS with coalescent simulations under specific demographic models through approximating likelihood is a good strategy to study genetic history [28,39]. We aimed to test whether DCs originated from *G. g. spadiceus* or from multiple genetic resources by explicitly accounting for demographic scenarios. Here, we presumed and tested three most likely models: DCs originated from multiple RJF genetic resources; DCs originated only from *G. g. spadiceus*; and DCs originated only from other RJF subspecies but not *G. g. spadiceus.* As we can clearly infer from above analyses (i.e., f_3 and f_4 statistics, TreeMix and qpGraph) that there is strong evidence of the introgression from one into another RJF subspecies when they have overlapped distribution ranges or from a particular RJF subspecies into a few local DC populations which are kept in the proximity to the RJF subspecies, in all these models, we considered asymmetric migrations between specific RJF subspecies and between particular RJF and relevant DC populations. All scenarios are shown in Supplementary information, Fig. S39. All computations were performed using fastsimcoal2 program [40]. Due to computational constraints, we simplified model tests by running these tests in three different data sets containing four populations. Population combinations are defined as follows:

(1) *G. g. murghi*, local DCs (Indian chickens), *G. g. spadiceus* and Yunnan chickens

(2) *G. g. gallus*, local DCs (Thai chickens), *G. g. spadiceus* and Yunnan chickens

(3) *G. g. jabouillei*, local DCs (Silkie and Vietnamese chickens), *G. g. spadiceus* and Yunnan chickens.

To mitigate the effect of linkage disequilibrium on our analysis, we took one SNP at least with 10 kb distance to generate SFS for each population. The unfolded SFS was generated using a modified script from *dadi* [41], which was the one we used before [6]. Estimations were performed with the following parameters: "-n 100000 -N 100000 -d -M 0.001 -l 10 -L40 -q -c 15". For each model and for running every data set/combination, 100 independent repeat runs with varying starting points were performed to ensure convergence. The results for all scenarios are shown in Supplementary information, Fig. S40.

7 Genome-wide scanning for variants introgressed from other three junglefowl species into DCs

7.1 *ZrIBD*

To demonstrate the presence of introgressed genomic regions from the other three species of the Green, Grey, and Ceylon Junglefowls into DCs, we leveraged similar methods as used by Bosse et al. [42] in identifying introgressed genomic tracts between European and Asian pigs. We also successfully used this method to explore the adaptive introgression among bovine species [43]. This approach has the advantage that it does not require an outgroup species to detect such admixture (e.g., we used Green Junglefowls as the outgroup for *D*- and *f*-statistic-based analyses). Firstly, we used Beagle software (version: 4.1) to impute and phase the genotypes [27]. Due to a lack of a phased panel for chicken currently, as stated [44], increasing the sample size and reducing the level of alleles with low frequency could improve Beagle's performance. In our study, alleles with frequency less than 5% were removed. Variants for all samples together were used to perform Beagle analysis. Secondly, shared haplotypes were counted in each pair of individuals. To quantify the shared haplotypes between the junglefowls and DCs across their whole genomes, we defined a topology presented in Supplementary information, Fig. S49.

Where A corresponds to each of the other three junglefowl species; B is donor DC, C is *G. g. spadiceus* from which B directly originated. We computed frequency of haplotypes (fIBD) shared for B and C and for B and A. The relative shared IBD tract between B and A relative to C (A and C are two competing groups) is defined as:

$$rIBD(B; C, A) = fIBD(B\text{-}C) - fIBD(B\text{-}A)$$

Where fIBD(B-C) = cIBD(B-C) / tIBD(B-C), of which cIBD(B-C) indicates the count of all haplotype IBD between groups of B and C, while tIBD(B-C) presents total number of all pairwise comparisons between groups of B and C; ranging from 0 (no IBD tract detected) to 1 (all individuals from B sharing IBD with all individuals from C). fIBD(B-A) was calculated in a similar manner. Sliding-window analyses were used to count haplotype frequency for each bin (window size of 10 kb with 5 kb increment). As the total number of pairwise comparisons differed between the groups, these numbers were normalized. The rIBD was Z-transformed as follows:

$$ZrIBD = (rIBD - \mu) / \sigma rIBD$$

Where μ is genome-wide mean for rIBD and σ is standard deviation. Empirically determined threshold for extreme IBD tract shared by B from A compared with C was set to 2 s.d. from the mean in the far-right tail of the distribution in different regions of the genome.

To test the power of this method, we used this method to excavate possible introgressed regions between Tibetan mastiff and Tibetan wolf. Using published genomes as documented in previous studies [45-48], including 16 Tibetan Mastiffs (TM), four Tibetan grey wolves (TW) and 32 Lowland grey wolves (LW), we computed rIBD (TM; TW, LW). To our expectation, the previously reported *EPAS1*, transferred from TW to TM via interbreeding to facilitate the latter to survive at high altitude [49,50], showed the strongest signature of introgression, implying the high sensitivity of our pipeline (Supplementary information, Fig. S49).

Here, we retrieved the junglefowl IBD tracts present in all DCs (DC) and local domestic chickens (LDC), as computed in the following combinations:

1): ZrIBD (Green Junglefowl; DC / LDC, *G. g. spadiceus*)

2): ZrIBD (Grey Junglefowl; DC / LDC, *G. g. spadiceus*)

3): ZrIBD (Ceylon Junglefowl; DC / LDC, *G. g. spadiceus*).

7.2 Testing for the direction of introgression using phylogenic tree

Due to the inherent attributes of our methods, we could not define the direction of admixture. So we constructed phylogenic tree based on haplotypes of putatively introgressed regions using MEGA7 [51] to determine the direction of introgression. If a junglefowl species clusters within DC clade in the tree, an introgression from DCs into the junglefowl species would be inferred; on the other hand, clustering of DCs together with a given junglefowl species in the tree would suggest the introgression from this junglefowl to DCs.

7.3 Note for the result

Using a Z-core >2 as cutoff, we conclude that Indian, Sri Lankan, and Indonesian chickens possess only 2%, 0.8%, and 0.2% of admixture with the Grey, Ceylon and Green Junglefowls, respectively. The results are shown in Supplementary information, Figs. S50-52.

8. Detection of positively selected genes in the domestic chicken

8.1 Detection of selective sweeps

We used locus-specific branch length (LSBL) statistics [52] akin to population branch statistics (PBS) [53] by taking genotypes from three populations. LSBL in population A was calculated for each SNP using the following formula LSBL(A;B,C) =(F_{ST}(AB)+F_{ST}(AC)-F_{ST}(BC))/2 [52]. Weir and Cockerham's F_{ST} [54] for each locus in each pair of populations (A and B; A and C; and B and C) was calculated using vcftools [15]. Sliding window analysis was performed with a 50 kb window size and 25 kb step size. Given that chickens originated from *G. g. spadiceus* and other RJF subspecies also contributed a proportion of ancestry to local chickens, LSBL statistics were calculated for the combination of LSBL(chicken; *G. g. spadiceus*, *G. g. jabouillei*), and LSBL(chicken; *G. g. spadiceus*, *G. g. murghi*). *G. g. gallus* and *G. g. bankiva* were excluded in this combination due to a limited sample size. We also leveraged the vcftools program [15] to calculate the nucleotide diversity (π) [55] for domestic chickens and the direct wild ancestor, *G. g. spadiceus*, using a 50 kb step with a 25 kb increment ("--window-pi 50000 --window-pi-step 25000"). The level of genetic diversity for domestic chickens relative to that of *G. g. spadiceus* was measured using the following formula: π-ratio = $\pi_{G.g.spadiceus}/\pi_{chicken}$. If a genomic region experienced strong selection in domestic chickens, an extremely high π-ratio would be expected. To identify positively selected genes (PSGs), we Z-transformed the π-ratio and LSBL of each window and retrieved windows with $Z \geq 3.3$, corresponding to $P \leq 0.001$. These windows were considered candidate selective sweeps and were then annotated to specific genes using the Variant Effect Predictor [56]. To investigate the population history on π-ratio and LSBLs, we used msms software [57] to simulated 1538 ("-I 3 54 72 1412") and 1620 ("-I 3 136 72 1412") neutrally evolved sequences with a length of 100 kb were simulated 10,000 times. The recombination rate was 2.8cM/Mb, and the mutation rate and generation time were 1.91×10^{-9} substitutions per site per year and one year, respectively [30]. Then, custom python scripts (available upon request) were used to convert the .ms format into a .vcf format. LSBLs and π-ratios were calculated using the same pipeline mentioned above for these sequences. The statistical significance of the selection score calculated from simulated and observed data was measured using the randtest function in the ade4 R package.

8.2. Functional enrichment analysis

To obtain a global perspective on the functions of candidate PSGs, we used g:Profiler[58] to retrieve the functional enrichment terms, including Gene Ontology (GO) categories, KEGG pathways, and Human Phenotype Ontologies (HPOs). Additionally, STRING (v10)[59], a web-based resource and application for exploring protein-protein interaction (PPI) networks, was used to explore the biological role of PSGs in a network.

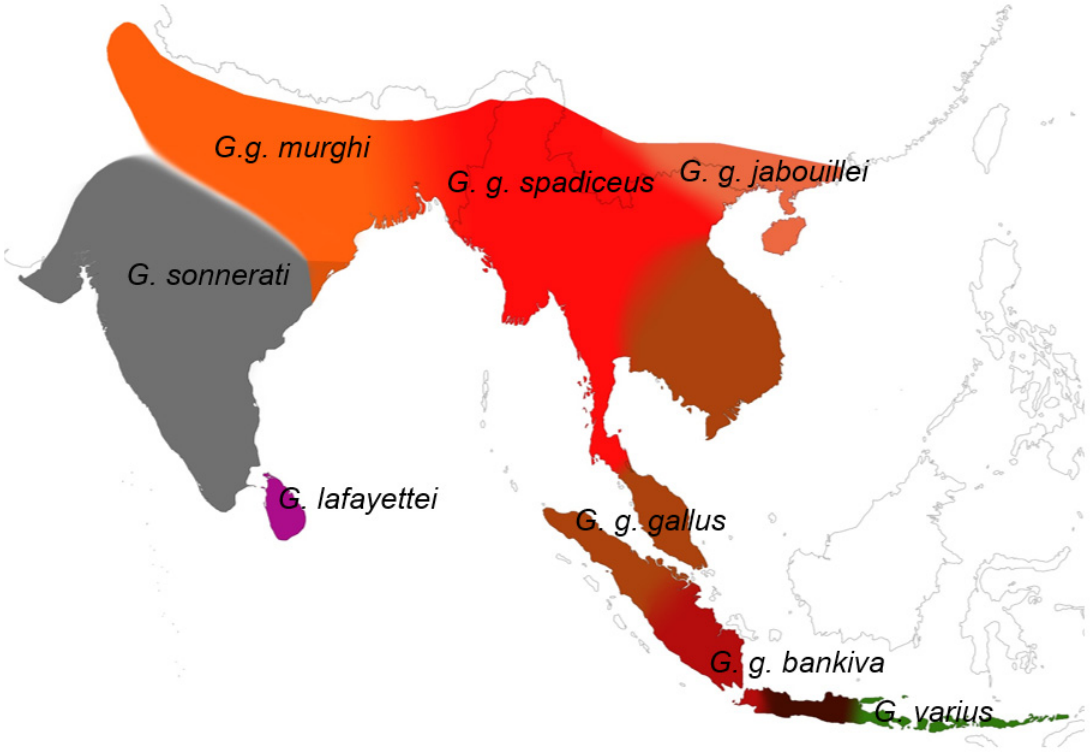
8.3 Supplementary Note for the result

We identified 176 ($LSBL_j$), 186 ($LSBL_m$) and 198 (π-ratio) putatively selected genes with high consistency based on our criteria ($Z \geq 3.3$), of which 163 and 86 genes were detected to be outliers by at least two and all three methods, respectively. Interestingly, all three statistics unequivocally support the strongest signature of selection in the chr22 region (Fig. 4a). The significance of these outliers, including the chr22 region, was further assessed by simulating neutrally evolving sequences. Our results suggest that, in all cases, drift alone is unlikely to explain these results (P<1e-4; Supplementary information, Fig. S53).

Similar to genomic surveys in other domestic animals [2,60-65], a number of genes showing signals of selection are associated with the nervous system, muscle and bone development and growth, and metabolic pathways (Supplementary information, Table S10-11). For instance, *IGF1, STC1, FGFR1, CHUK, CRMP2A, LCORL* and *NODAL* are involved in morphological changes (e.g., muscle and bone development, body size and growth regulation); *ADRA2C, APBA2, PRNP, BDNF PARD3, ADGRL3, NFEL, S100B* and *NEFM* are implicated in the regulation of nervous system and behavioral traits, such as aggressiveness; *CAMK2D* and *ADCY1* are associated with melanogenesis. Specifically, *IGF1* (Z_{LSBLm}=4.2; Z_{LSBLj}=3.67; and $Z_{\pi\text{-ratio}}$=1.47), encoding insulin-like growth factor 1 involved in growth regulation, plays a substantial role in controlling body size across different dog and chicken breeds [6,66]. *ADRA2C* (Z_{LSBLm}=3.8; Z_{LSBLj}=4.0; and $Z_{\pi\text{-ratio}}$=4.1), an inhibitory modulator of the sympathetic nervous system, is involved in fight-or-flight response. Knocking out this gene in mice resulted in significant behavioral changes, such as enhanced startle response, shortened attack latency and diminished acoustic prepulse inhibition, whereas its overexpression led to opposite changes [67]. *CAMK2D* (Z_{LSBLm}=5.12; Z_{LSBLj}=4.47; and $Z_{\pi\text{-ratio}}$=3.5) encodes an enzyme that is a member of the calcium/calmodulin-dependent protein kinase II family. *CAMK2D* is known for its role in

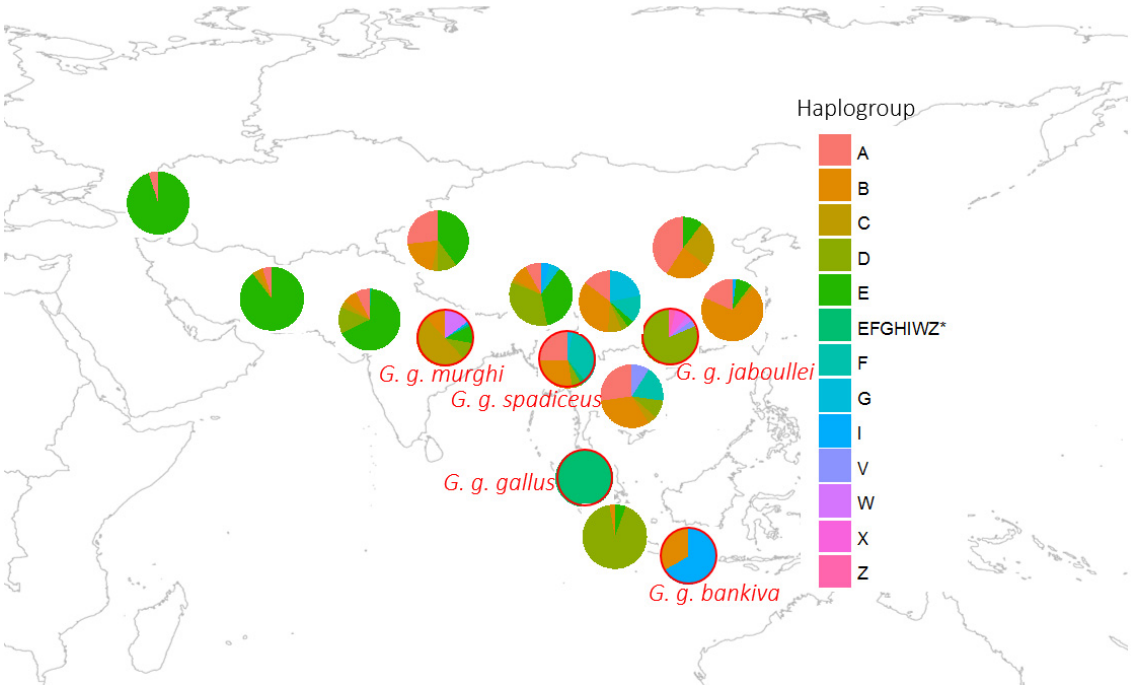
synaptic plasticity and memory formation [68]. All these three genes have also been previously reported as exhibiting signals of selection in domestic hens [6,12,69], highlighting their potential roles in chickens and reflecting the effectiveness of the statistics we used here.

Supplementary Figures



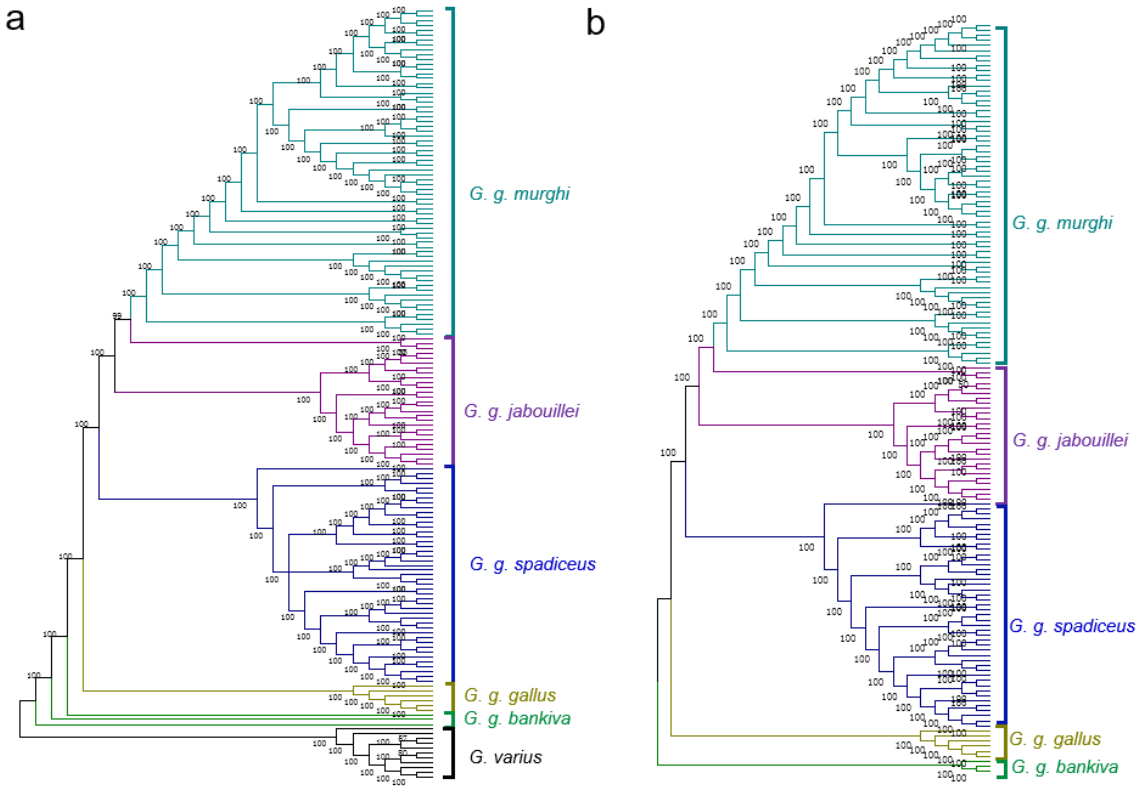
Supplementary information, Fig. S1.

Map showing the geographic distribution ranges of all wild junglefowl species across South Asia, Southeast Asia, and East Asia. For Red Junglefowls, *G. g. murghi* is found in northern Indian subcontinent; *G. g. spadiceus* in Mainland Southeast Asia (MSEA, Yunnan province of southwest China, western Thailand and Myanmar); *G. g. gallus* in Island Southeast Asia (ISEA; Sumatra of Indonesia and the Philippines); *G. g. jabouillei* in southern China (Guangxi and Hainan provinces) and Vietnam; and *G. g. bankiva* in southern Sumatra, Java, and neighboring islands of Indonesia. All the information for the distribution ranges of extant wild junglefowl species is retrieved from relevant references [70-75].



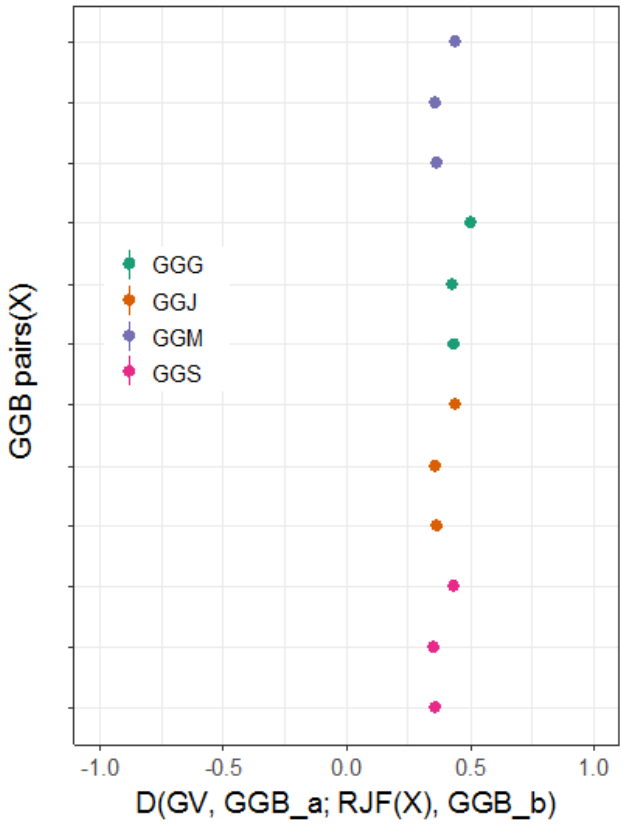
Supplementary information, Fig. S2.

Geographic distribution of mitochondrial DNA haplotypes of Red Junglefowls and domestic chickens. Haplotype proportions are shown in pie plots. Pie with and without red borders represent the proportions of different haplotypes identified in Red Junglefowls and domestic chickens, respectively. Mitochondrial DNA from each sample was assigned into specific haplotypes using DomeTree program [37].



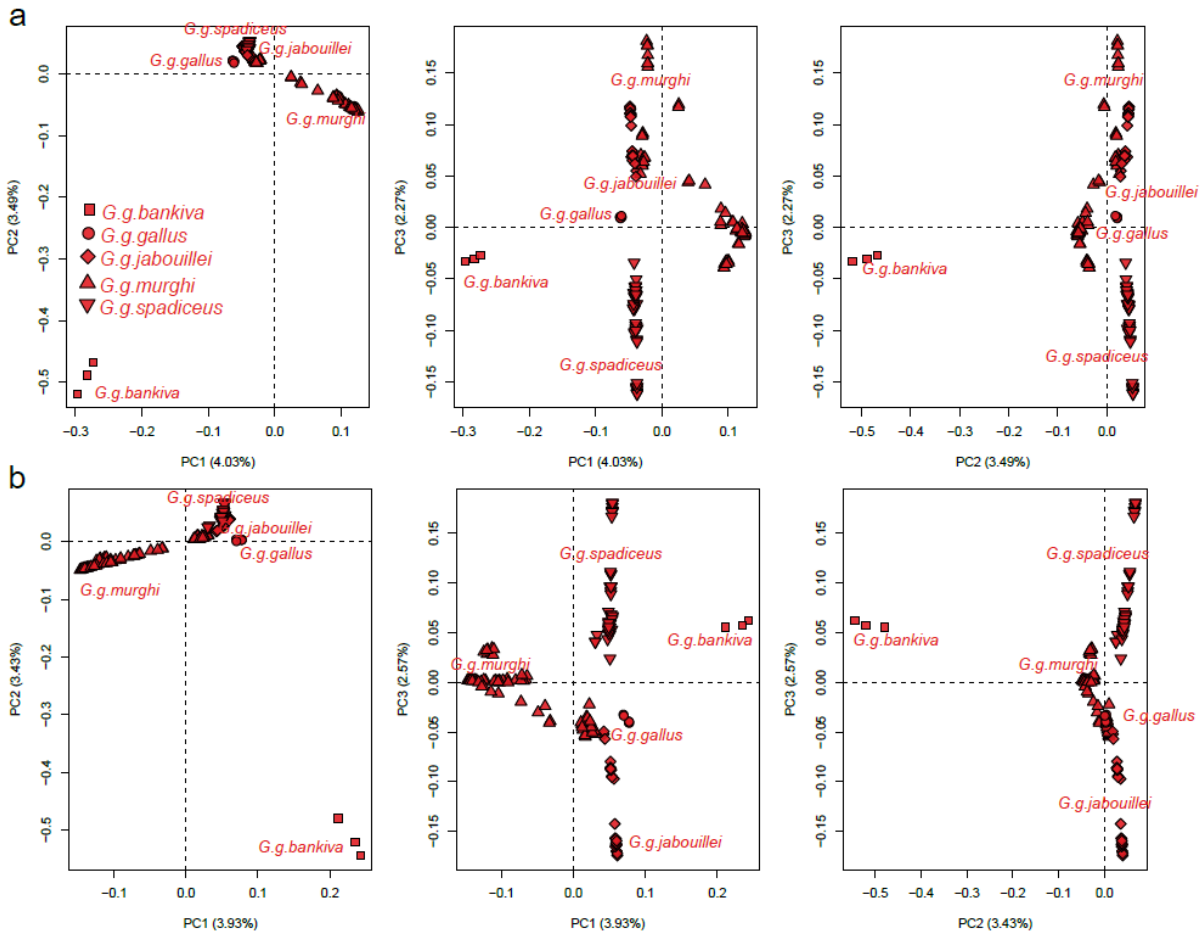
Supplementary information, Fig. S3.

Maximum-likelihood phylogenetic trees constructed using FastTree program (version 2.1.9; available at http://www.microbesonline.org/fasttree/) based on whole autosomal genome data set. **a** Tree was constructed using genomes from all Red Junglefowls and rooted using the Green Junglefowl. The genetic relationship for all wild junglefowls reported here and in Fig. 1b is consistent with the mitochondrial DNA tree [76]. **b** Tree was constructed using the genomes from all Red Junglefowl subspecies and *G. g. bankiva* was used to root the tree. For construction of these trees, 100 bootstraps were analyzed, and bootstrap values are indicated by numbers at the base of the nodes. These analyses indicate that all Red Junglefowls cluster into five distinct lineages according to their subspecies classification except for three *G. g. jabouillei* individuals clustered within *G. g. murghi* lineage, due likely to their admixture given that the distribution ranges of *G. g. jabouillei* and *G. g. murghi* are overlapped with *G. g. spadiceus*, however, incomplete lineage sorting and sequencing artifacts could also affect such topology.



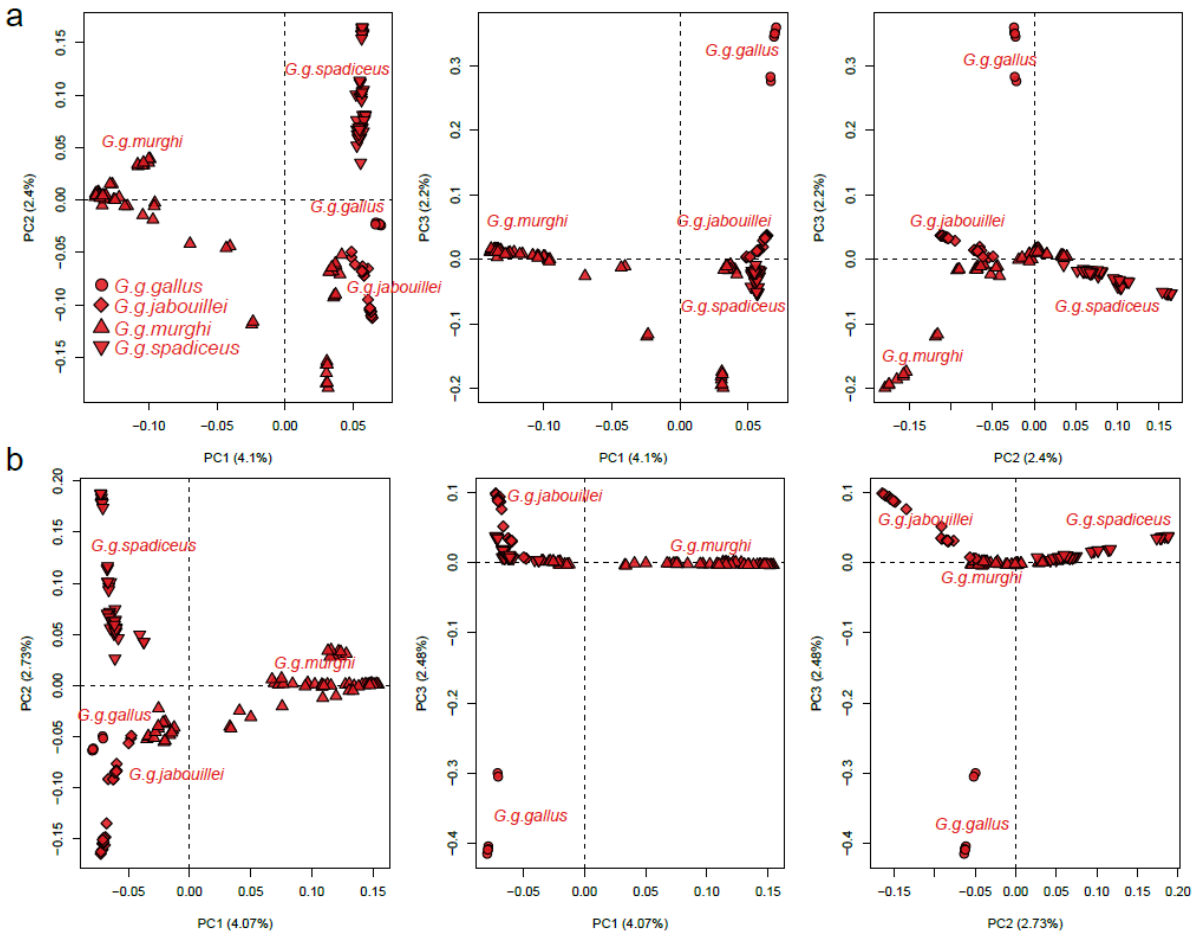
Supplementary information, Fig. S4.

This figure represents the *D*-statistic to assess whether the two groups of *G. g. bankiva* (GGB) in Supplementary information, Fig. S3a is more closely related to each other than to other Red Junglefowl subspecies. The result based on *D*(GV, GGB_a; Red Junglefowl, GGB_b) > 0, indicates that the two groups (GGB_a and GGB_b) are more closely related to each other than to any other Red Junglefowl subspecies. *D*-statistics used here were computed using qpDstat program of ADMIXTOOLS [22].



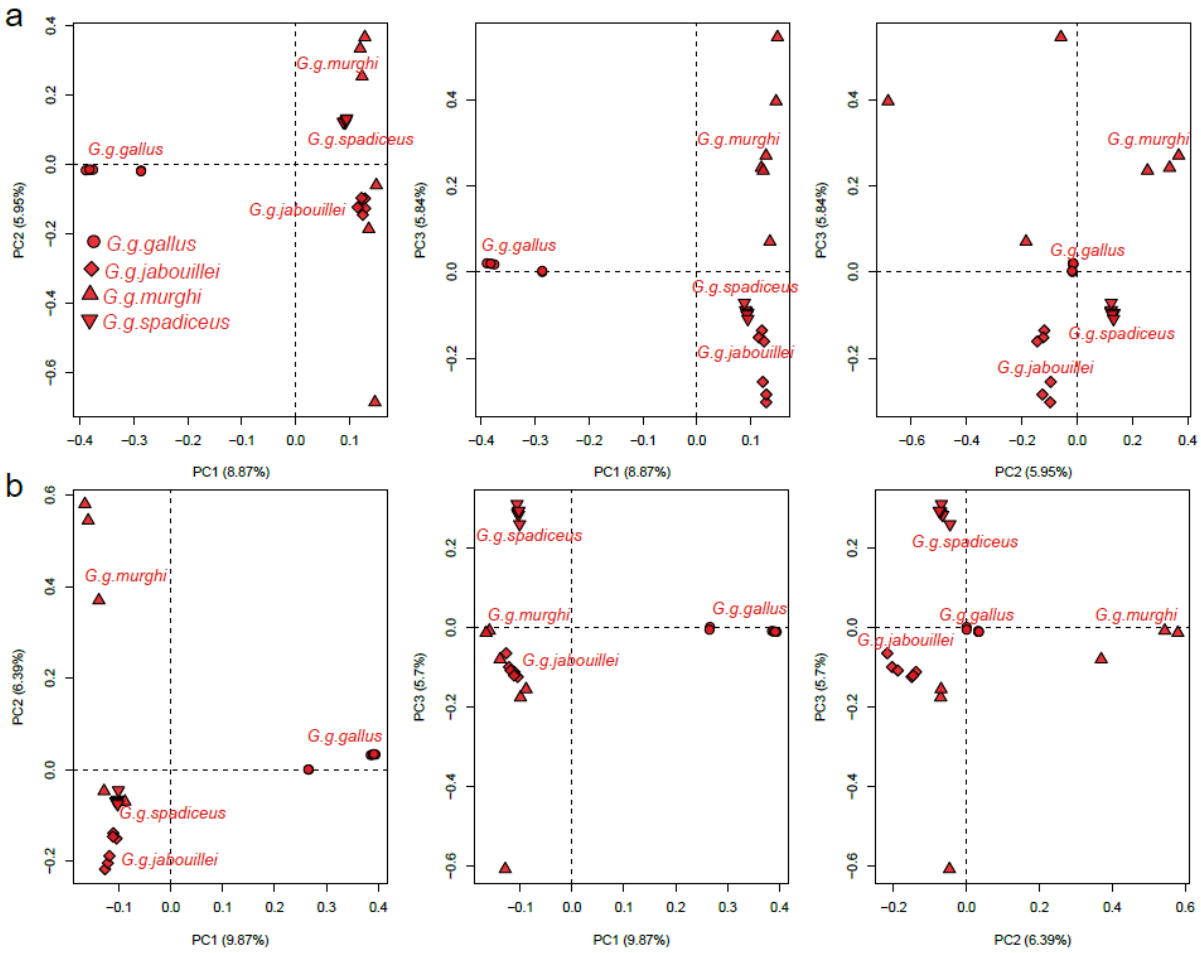
Supplementary information, Fig. S5.

Principal component (PC) analysis performed using GCTA software [16] (**a**) and smartpca program [17] (**b**) indicates the stratification of Red Junglefowl subspecies. The analysis was based on the autosomal variants pruned by PLINK toolset [77] with the following parameters "--indep 100 50 0.2". The genetic cline is seen in the plot of PC1 against PC3 for *G. g. murghi* across its geographical range.



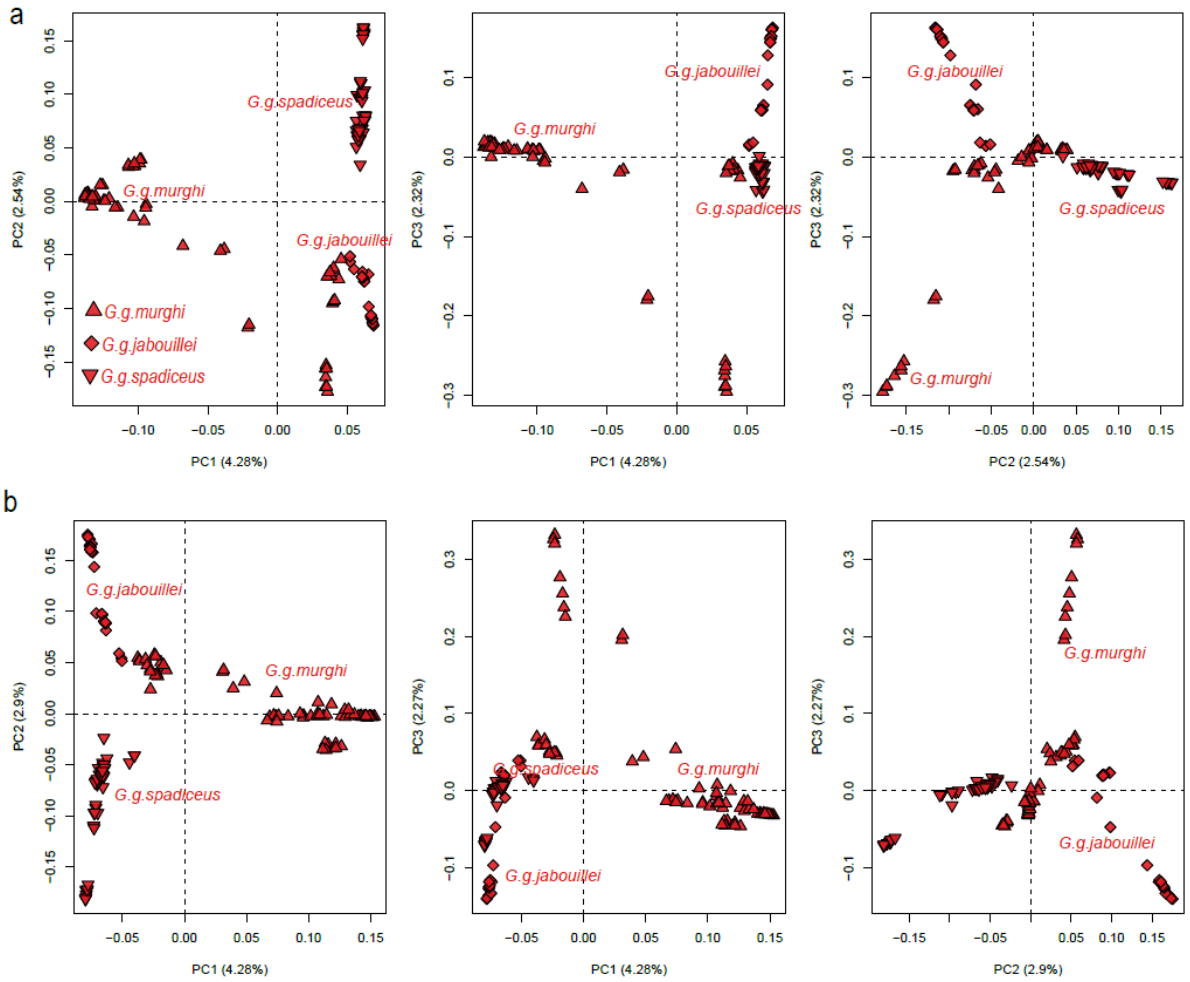
Supplementary information, Fig. S6.

Principal component (PC) analysis performed using GCTA software [16] (**a**) and smartpca program [17] (**b**) for all Red Junglefowl subspecies by excluding *G. g. bankiva* due to its high divergence from with other four Red Junglefowl subspecies. The analysis was based on the autosomal variants pruned by PLINK toolset [77] with the following parameters "--indep 100 50 0.2". The genetic cline is seen in all three plots for *G. g. murghi* across its geographical range.



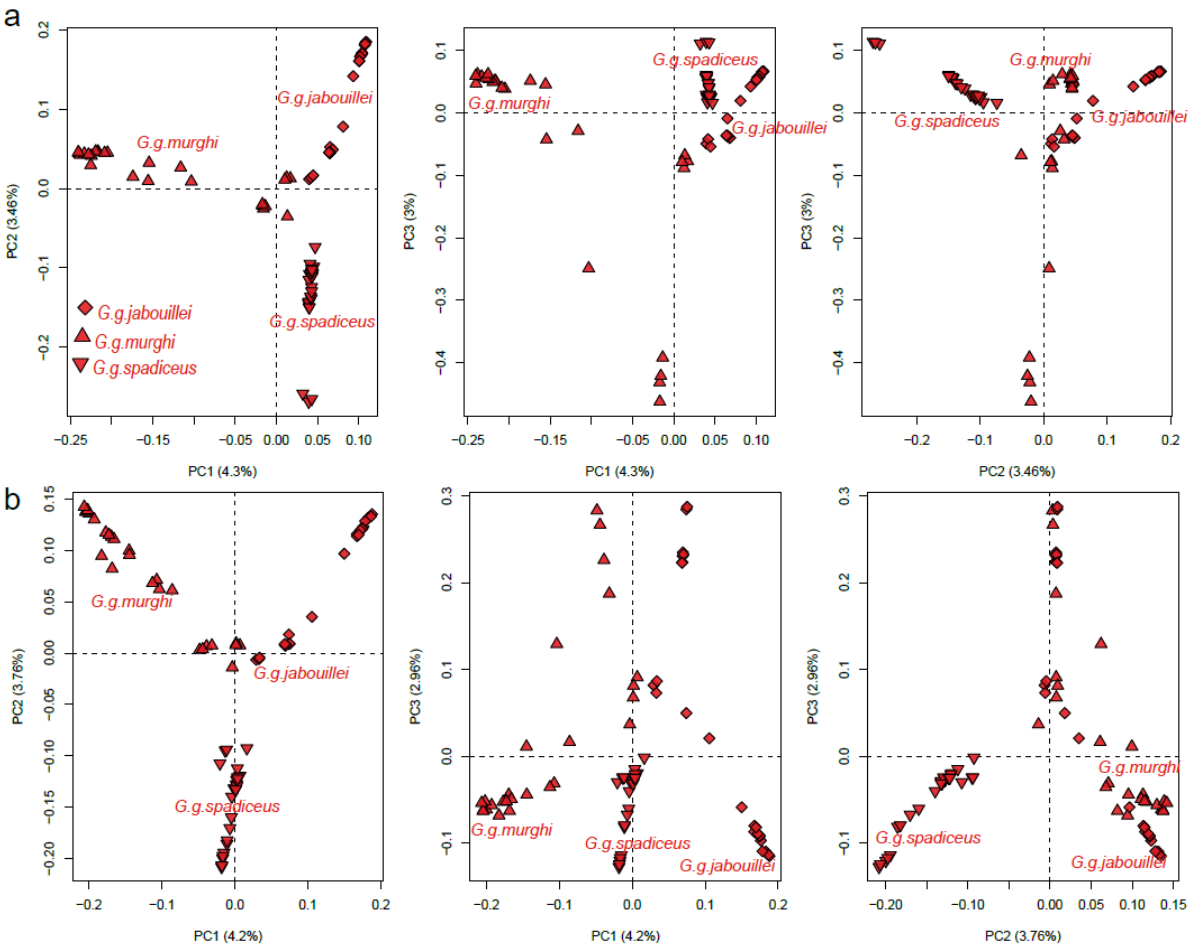
Supplementary information, Fig. S7.

Principal component (PC) analysis performed using GCTA software [16] (**a**) and smartpca program [17] (**b**) for all Red Junglefowl subspecies by excluding *G. g. bankiva* due to its high divergence from with other four Red Junglefowl subspecies. Six samples for each RJF subspecies were randomly selected for this analysis. The analysis was based on the autosomal variants pruned by PLINK toolset [77] with the following parameters "--indep 100 50 0.2". The genetic cline is seen in all three plots for *G. g. murghi* across its geographical range.



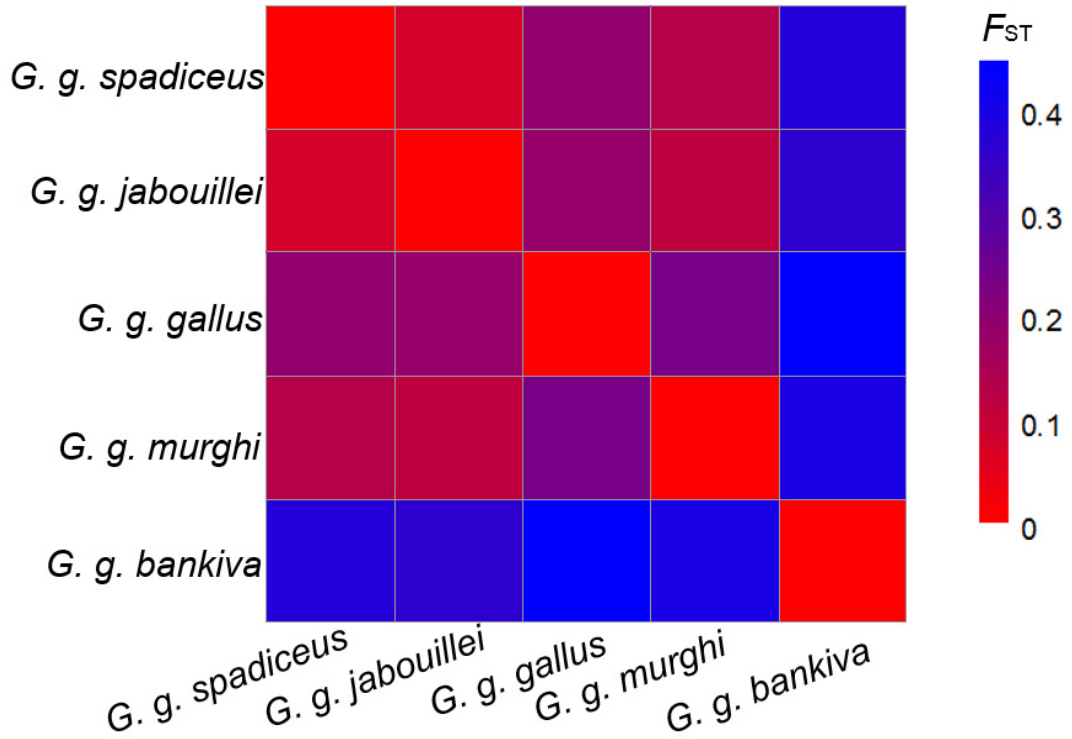
Supplementary information, Fig. S8.

Principal component (PC) analysis performed using GCTA software [16] (**a**) and smartpca program [17] (**b**) for all *G. g. murghi, G. g. jabouillei* and *G. g. spadiceus*.



Supplementary information, Fig. S9.

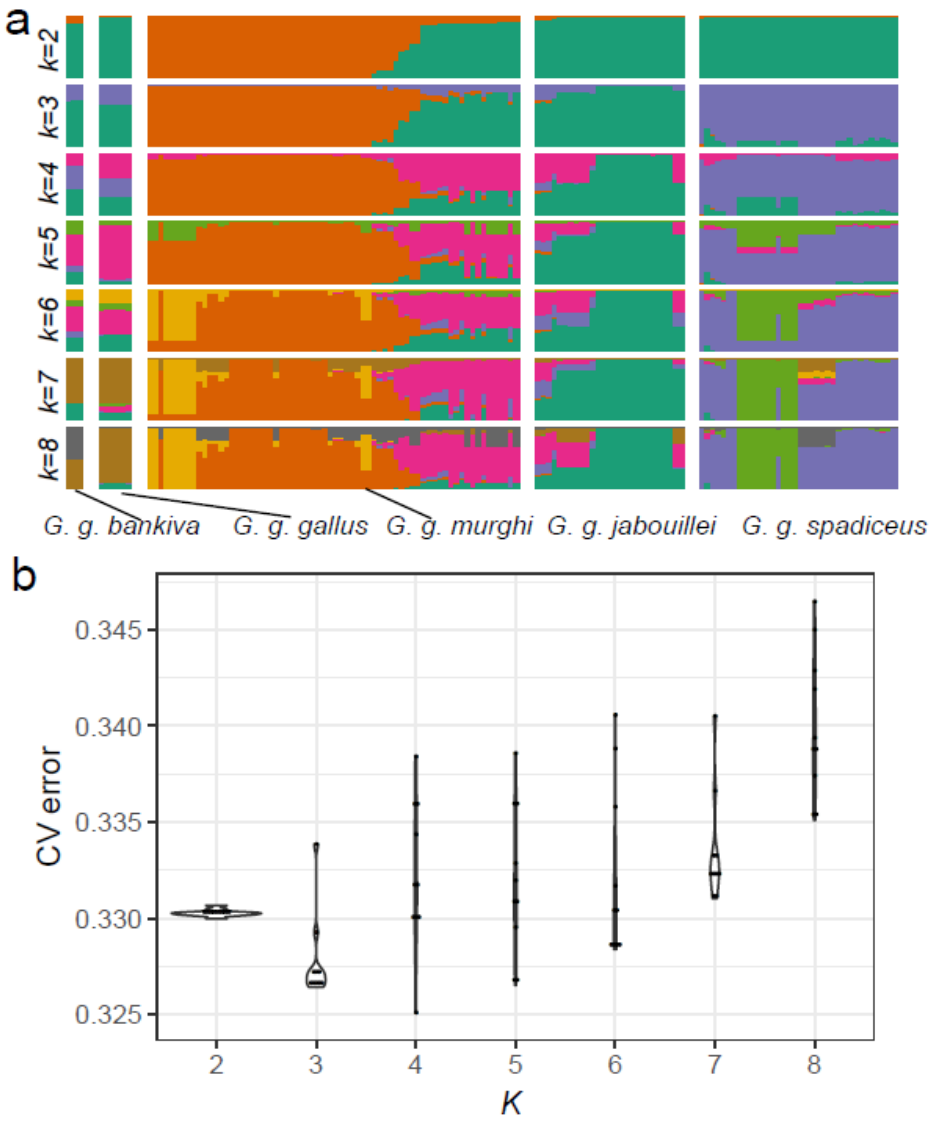
Principal component (PC) analysis performed using GCTA software [16] and smartpca program [17] for 26 random selected samples from each of *G. g. murghi, G. g. jabouillei* and *G. g. spadiceus*.



Supplementary information, Fig. S10.

F_{ST} matrix computed for each pair of Red Junglefowl subspecies.

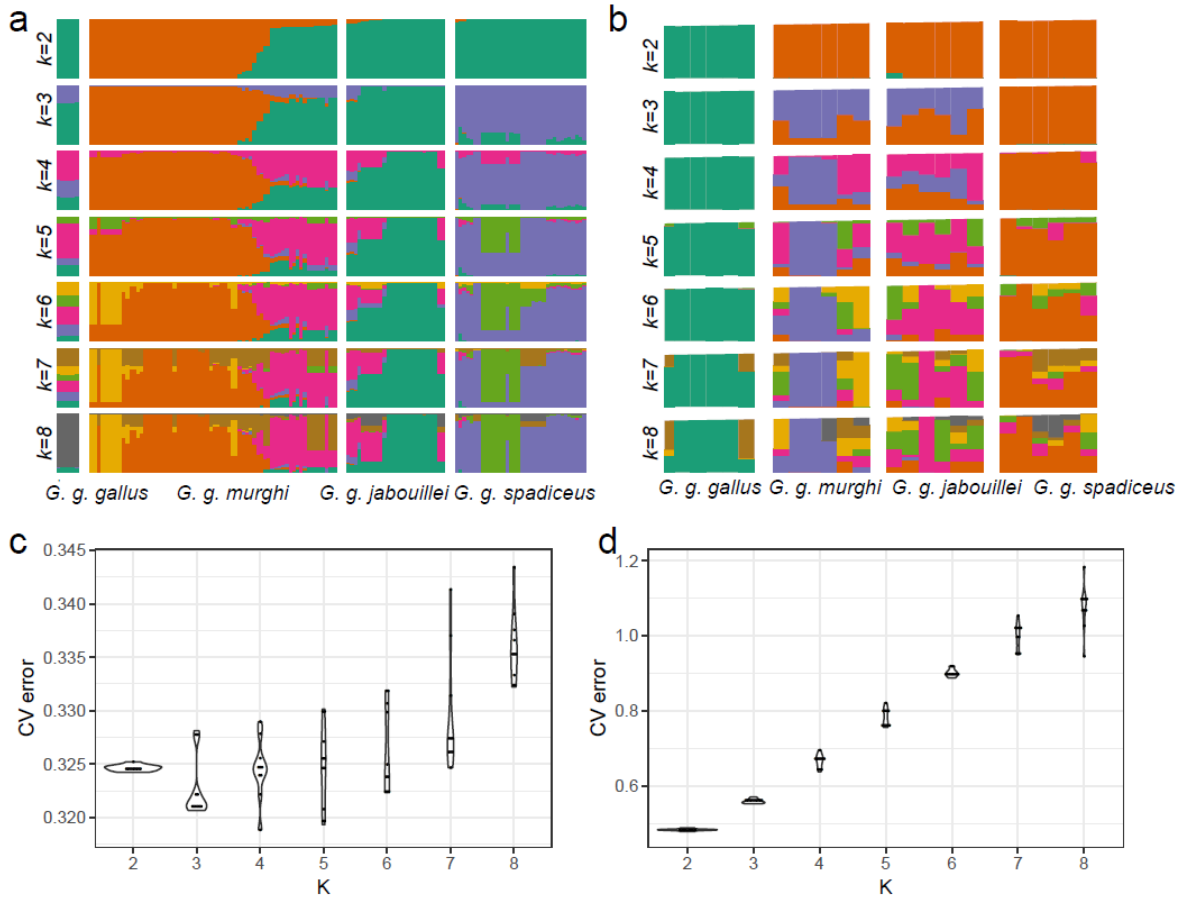
5



Supplementary information, Fig. S11.

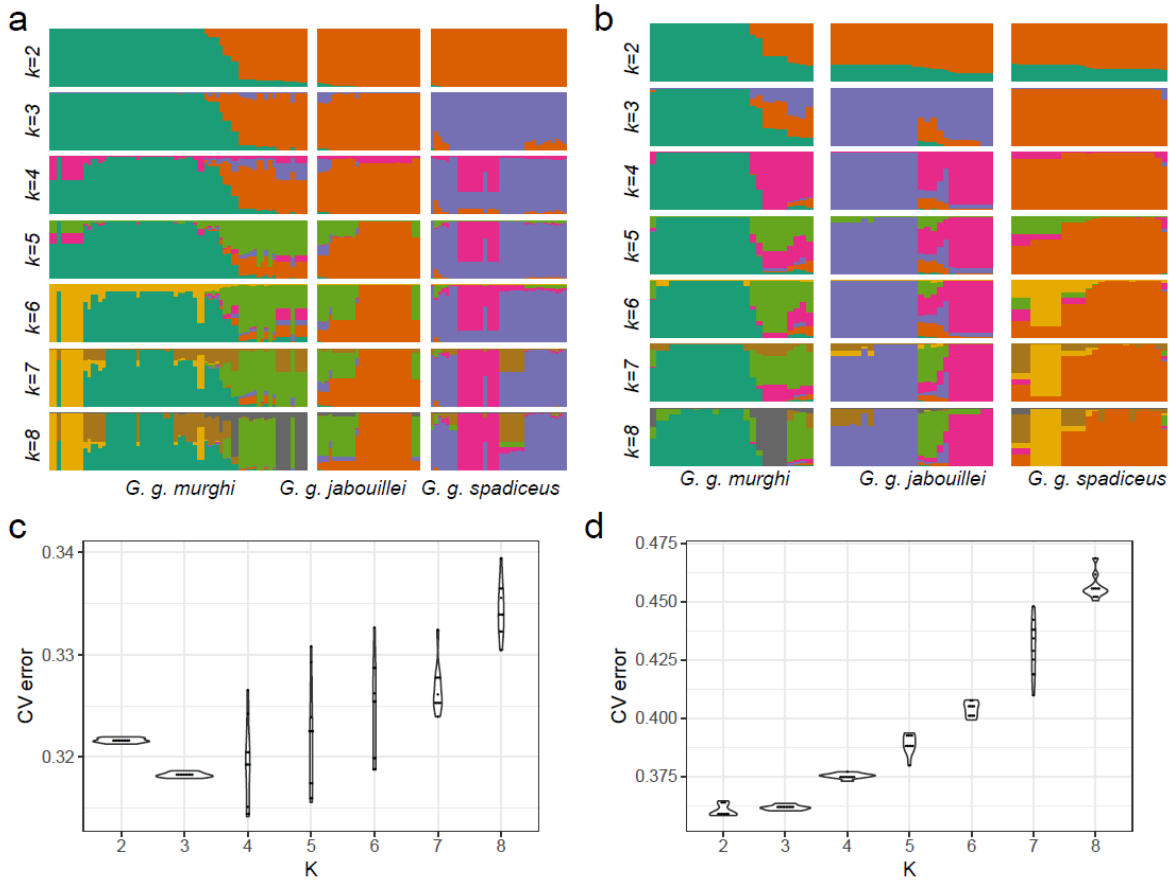
a The admixture bar plots (denoting K = 2 to 8 from top to bottom) showing the population stratification in Red Junglefowl subspecies. The analysis was based on the autosomal variants pruned by PLINK toolset [77] with the following parameters "--indep 100 50 0.2". The genetic cline is seen for *G. g. murghi* across its geographical range from northwest to northeast of India.

b The CV error for the ADMIXTURE analysis at K values ranging from 2 to 8.



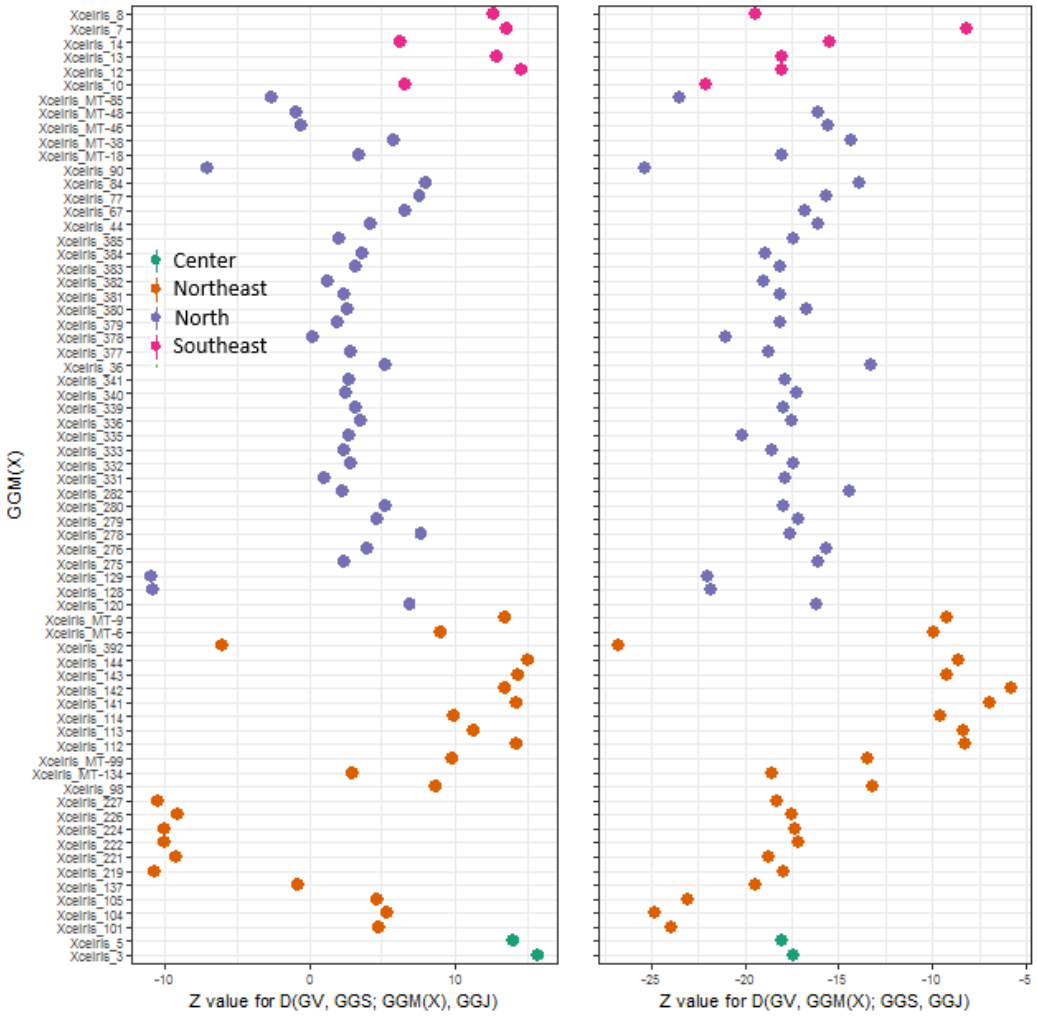
Supplementary information, Fig. S12.

The admixture bar plots (denoting $K = 2$ to 8 from top to bottom) showing the ancestral components for each Red Junglefowl subspecies. *G. g. bankiva* was removed from the analysis due to its high divergence from with other four Red Junglefowl subspecies. **a, b** The analysis performed for all samples and six randomly selected samples for each subspecies, with CV error shown in **c** and **d**, respectively.



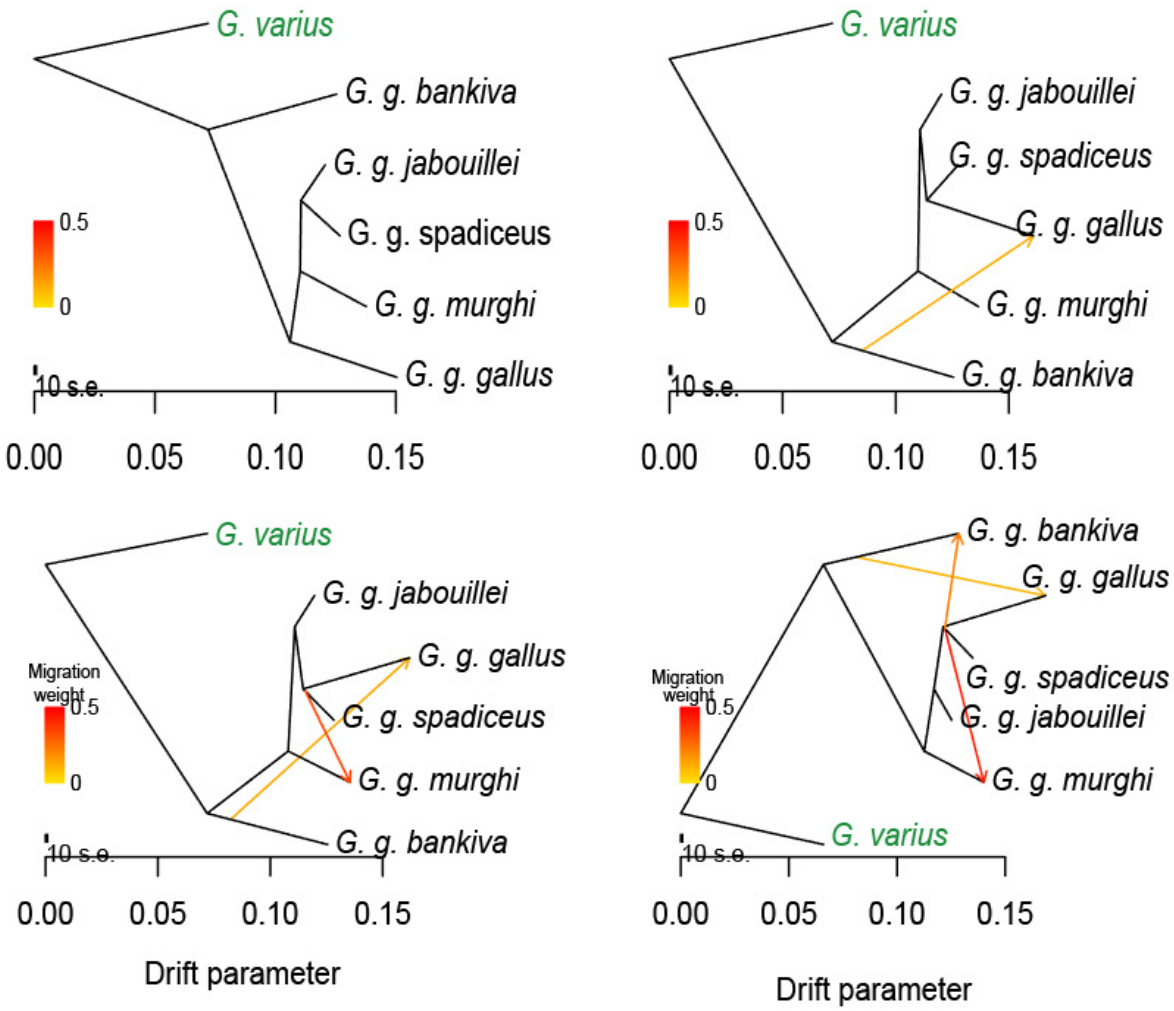
Supplementary information, Fig. S13.

The admixture bar plots (denoting K = 2 to 8 from top to bottom) showing the ancestral components for *G. g. murghi, G. g. jabouillei* and *G. g. spadiceus*. **a, b** The analysis performed for all samples and six randomly selected samples for each subspecies, with CV error shown in **c** and **d**, respectively.



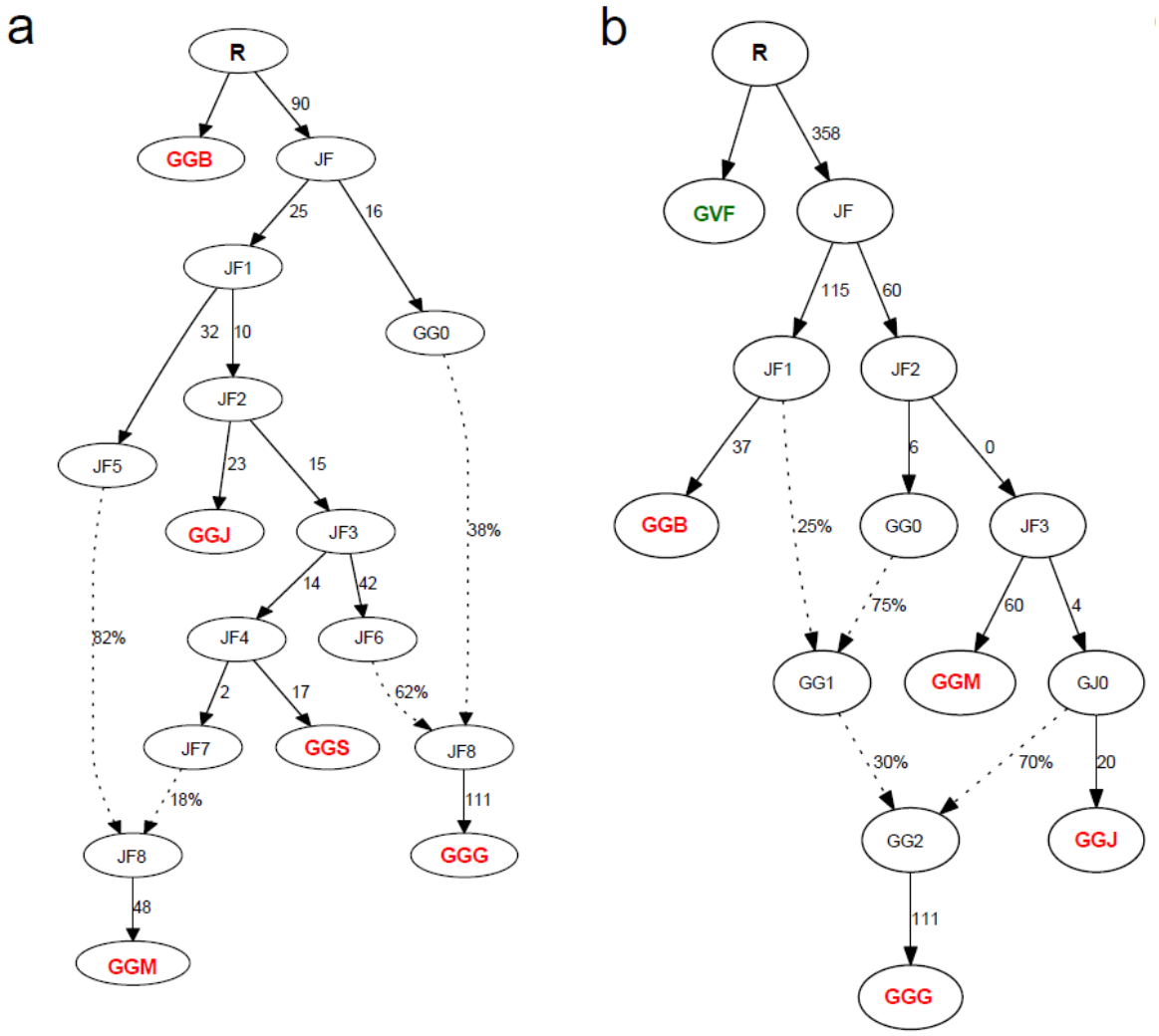
Supplementary information, Fig. S14.

The *D*-statistics reveal that several *G. g. murghi* (GGM) individuals from its easternmost distribution range in northern India (edge of Myanmar where *G. g. spadiceus* (GGS) is distributed) have closer relationships with *G. g. spadiceus* than those from other parts of its range, indicating interbreeding with *G. g. spadiceus*. The results also identify admixture between GGS and *G. g. jabouillei* (GGJ). The vertical axis shows each GGM individuals marked in colors according to their sampling locations within its range. Note: Birds of GGM collected from its southeastern distribution range were mostly from local zoos.



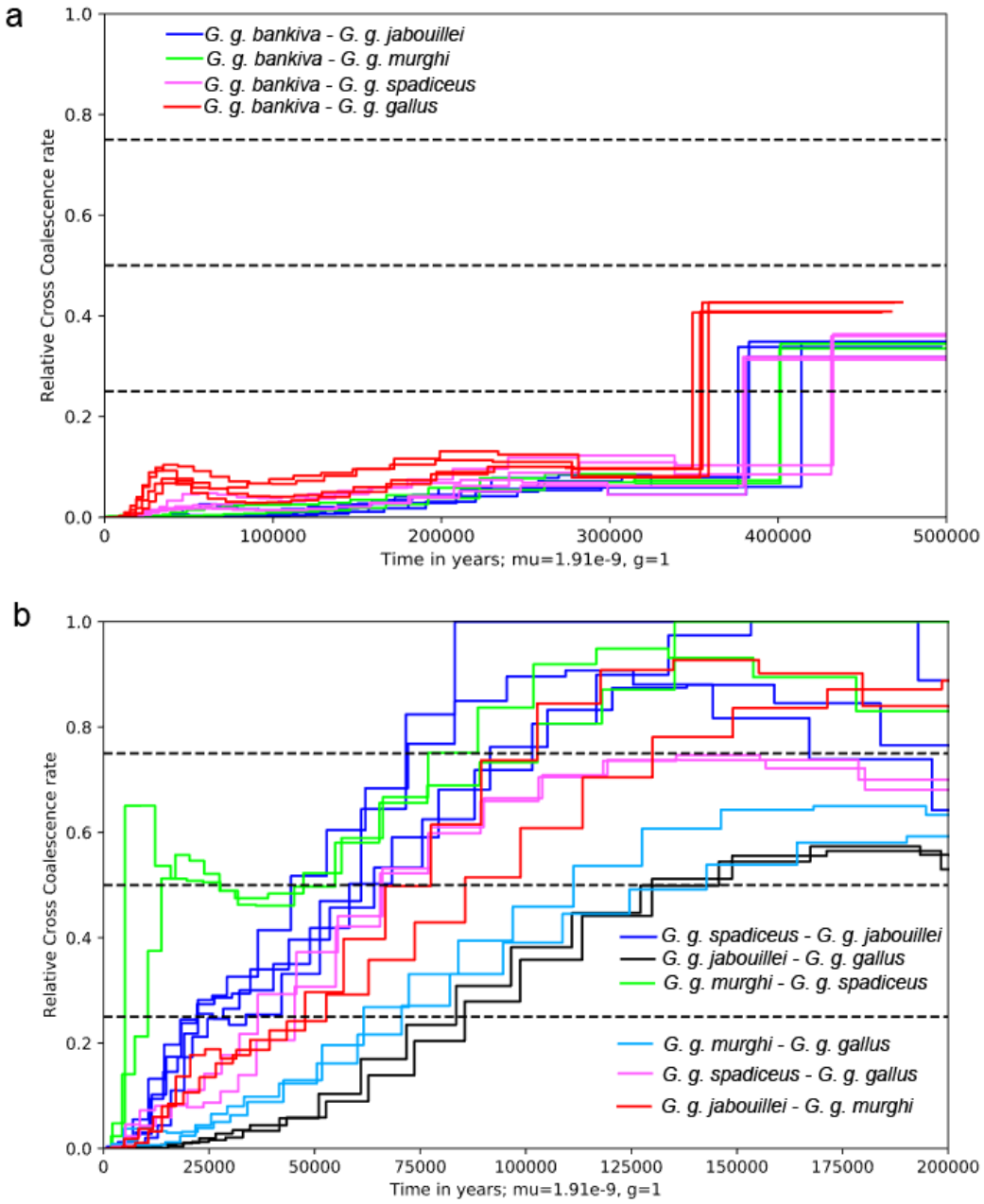
Supplementary information, Fig. S15.

The TreeMix analysis reveals admixture (from 0 to 3) and the splitting of Red Junglefowl subspecies. The earliest split occurs between *G. g. bankiva* and other four Red Junglefowl subspecies. Without migration (m = 0), this topology could explain > 99.964% of variance for the autosomal SNP data. Increasing migration events could marginally improve fitness of the model, with m = 1, 2, and 3 explaining 0.9999214, 0.9999747, and 0.9999996 of the variance, respectively. There are signatures of admixture between subspecies pairs with overlapped distribution ranges, such as between *G. g. bankiva* and *G. g. gallus* as well as between *G. g. murghi* and *G. g. spadiceus*.



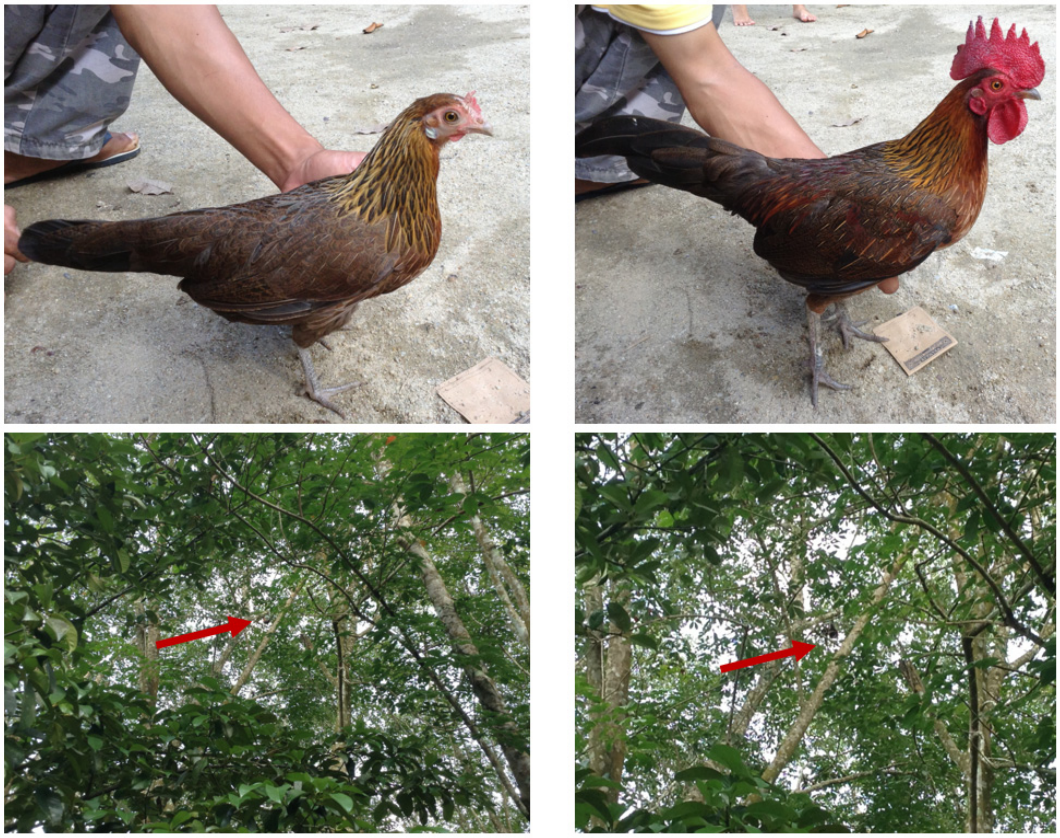
Supplementary information, Fig. S16.

a Best fitting model inferred by qpGraph program for *G. g. gallus* (GGG), *G. g. spadiceus* (GGS) and *G. g. jabouillei* (GGJ), with *G. g. bankiva* (GGB) using as outgroup (the most divergence lineage in PCA and phylogenetic analysis). **b** The best-fitted model inferred by qpGraph program suggests that there is a complex admixture among *G. g. gallus* (GGG), *G. g. bankiva* (GGB) and *G. g. jabouillei* (GGJ), where the Green Junglefowl (GV) is used as an outgroup. Branch lengths are shown in units of $F_{ST} \times 1000$ and dashed lines indicate inferred admixture events.



Supplementary information, Fig. S17.

Relative cross coalescence rate (CCR) obtained from MSMC. **a** This figure demonstrates that *G. g. bankiva* is the most divergent and separates from the other four Red Junglefowl subspecies > 500,000 years ago (based on CCR = 50%). It also shows that *G. g. gallus* has admixed with *G. g. bankiva* (peaks marked by red line). **b** This figure shows that the remaining four Red Junglefowl subspecies separated from each other around 50,000-125,000 years ago (CCR = 50%). The light green line shows some peaks of cross coalescence between *G. g. murghi* and *G. g. spadiceus,* indicating gene flow between these two subspecies took place within the past 25,000 years. These results also support that *G. g. bankiva* is the first diverged RJF subspecies followed by *G. g. gallus*, consistent with phylogenic analysis. We find that CCR curves computed for *G. g. gallus* with *G. g. jabouillei*, *G. g. murghi* and *G. g. spadiceus* do not reach to 1, which are due likely to the admixture of *G. g. gallus* with other Red Junglefowls in their overlapped ranges (Supplementary information, Figs. S14-16). This pattern is also tested using the simulations as described in references [45,46,78].

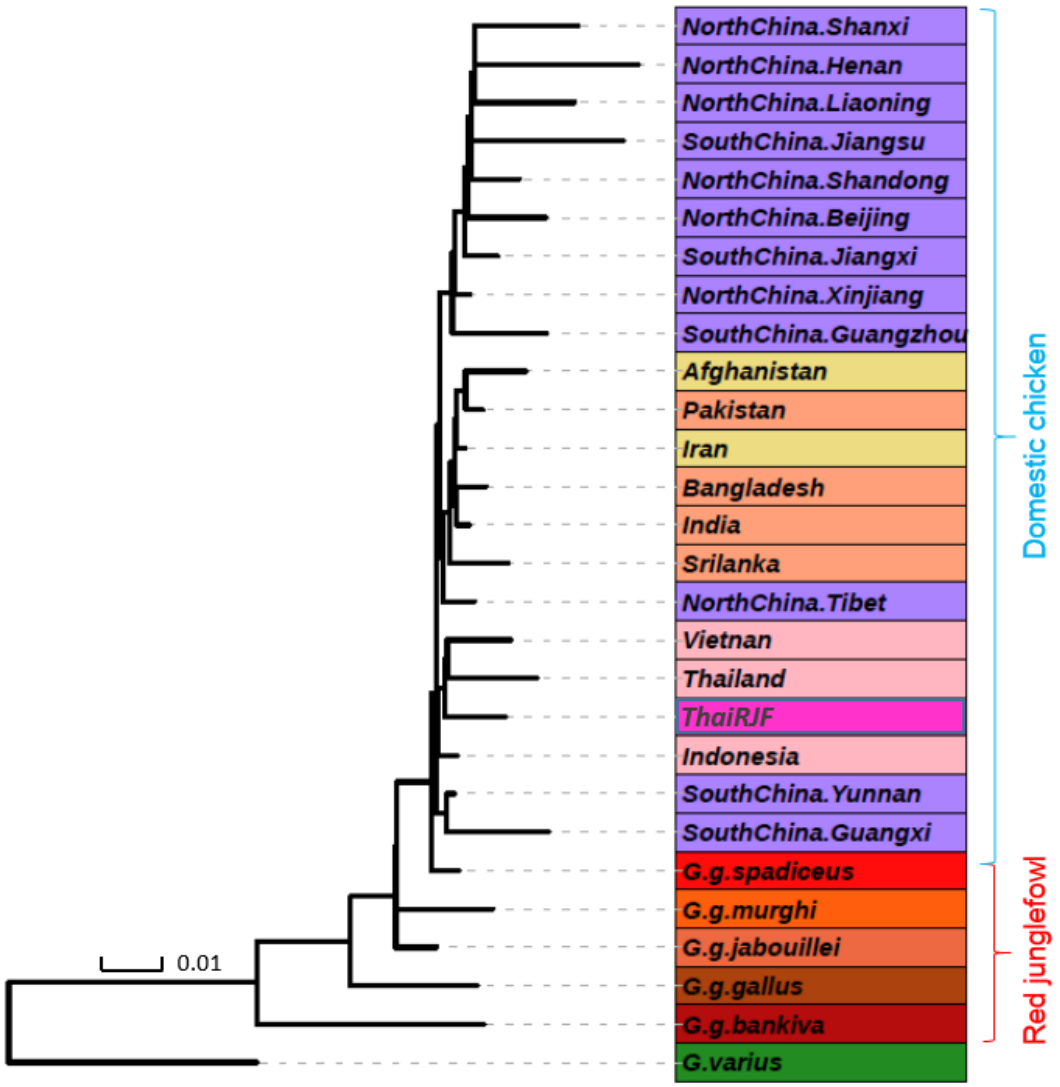


Supplementary information, Fig. S18.

Photos of the two indigenous chickens sampled from Thailand fall into the *G. g. spadiceus* lineage in the basal position of the phylogenetic tree (Supplementary information, Fig. S2a). By applying f_4-statistics, these two birds were identified to have higher genetic make-up of domestic chickens at the base of the tree (Supplementary information, Table S3, named as Thailand-2) than *G. g. spadiceus*.

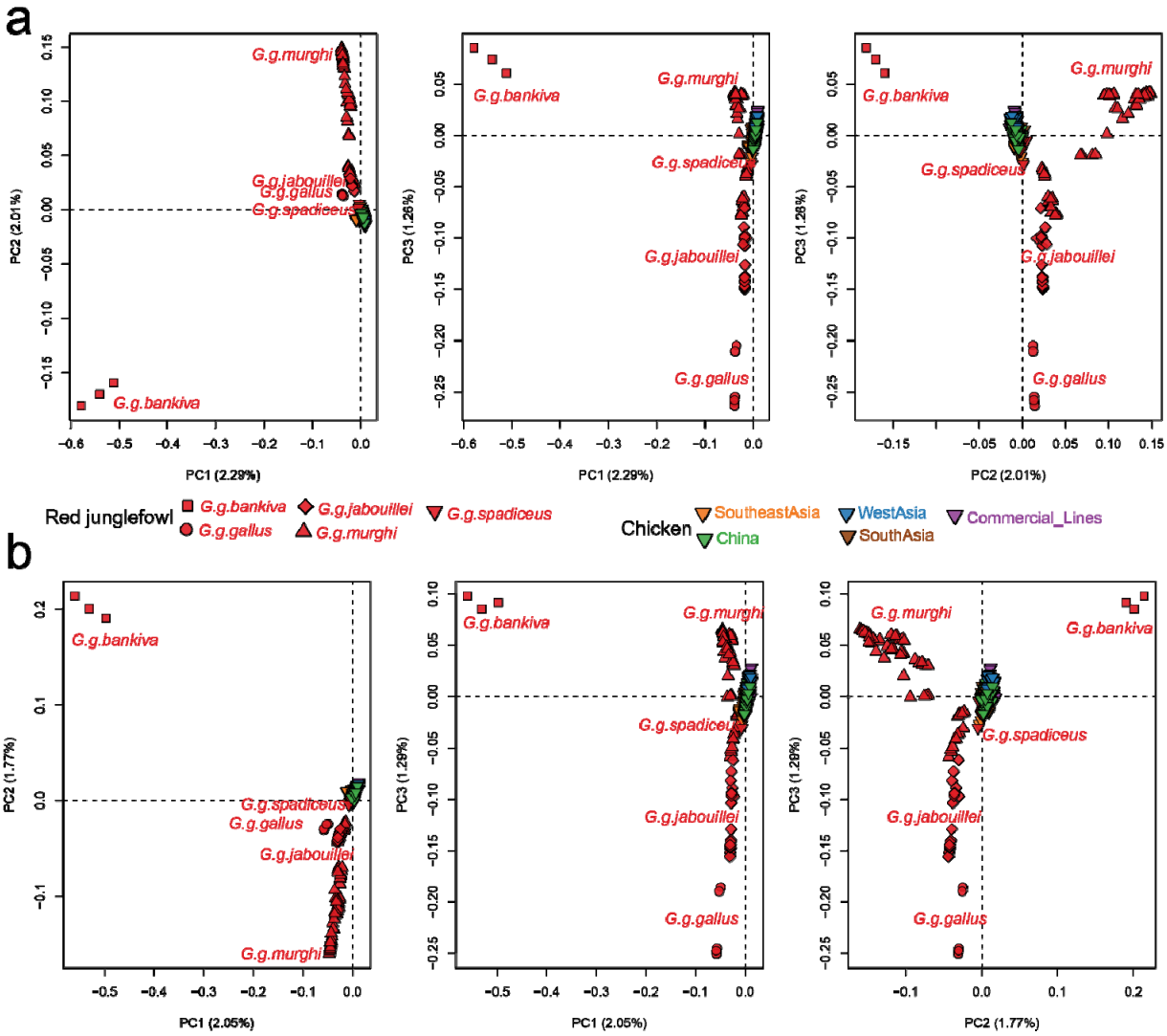
Supplementary information, Fig. S19.

Red Junglefowls of *G. g. spadiceus* sampled in Thailand (referred to as ThaiRJF) cluster with domestic chickens in phylogenic tree (Fig. 2a). They have likely admixed with indigenous chickens kept in their proximity (Supplementary information, Table S3).



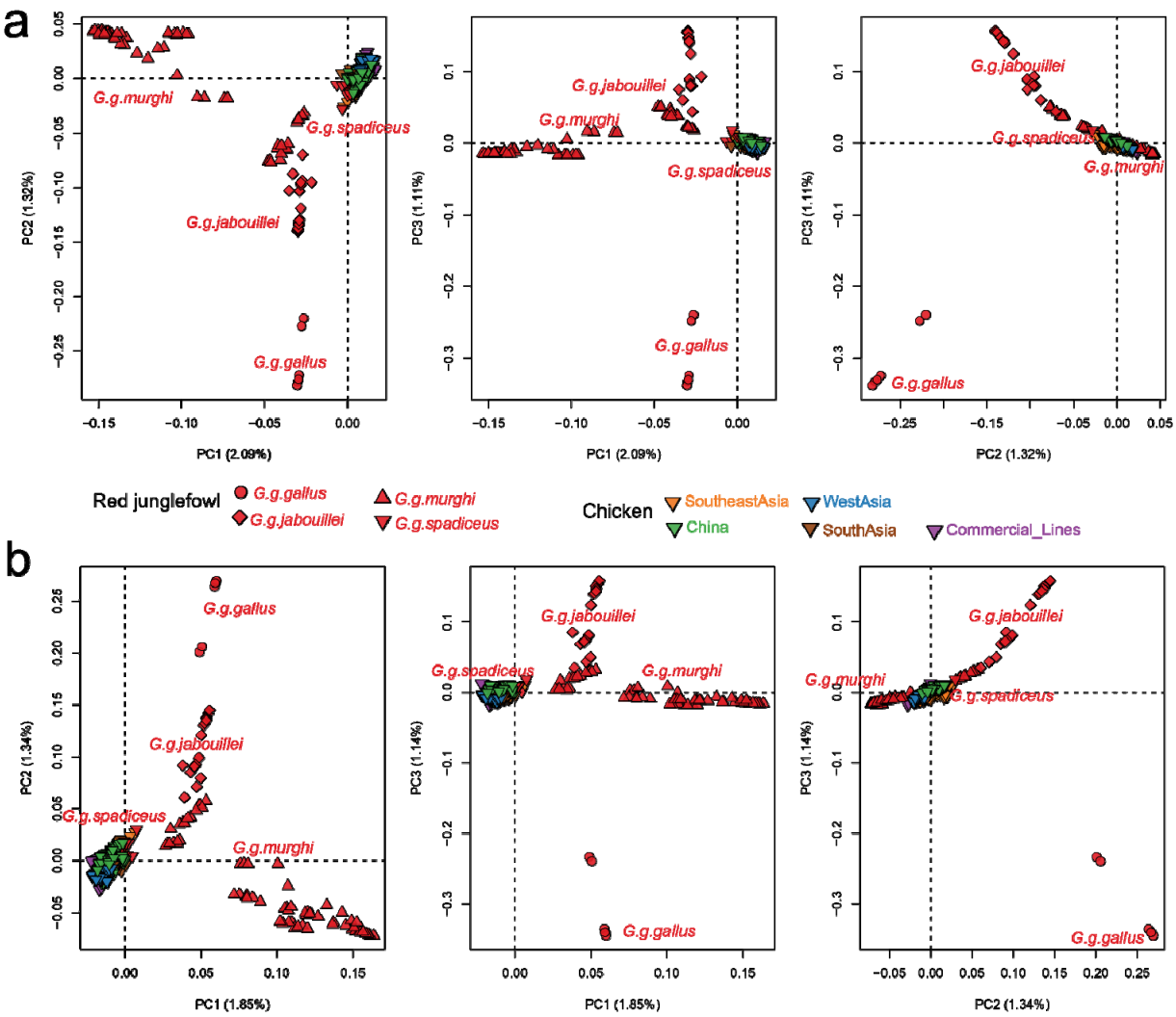
Supplementary information, Fig. S20.

Maximum-likelihood tree constructed with each subspecies (also a population of *G. g. spadiceus* collected from western Thailand, ThaiRJF) of Red Junglefowls and every population of domestic chickens as a unit using TreeMix program without migration edge (m = 0). The Green Junglefowl (*G. varius)* was used as an outgroup. The tree constructed with Red Junglefowls and domestic chickens excluding commercial birds. Indigenous chickens were grouped according to their sampling locations. In a close agreement with the tree based on individuals (Fig. 2a), Red Junglefowl populations also cluster according to their subspecies classification. *G. g. spadiceus* samples obtained from western Thailand (ThaiRJF) cluster within the domestic chicken lineage while the remaining *G. g. spadiceus* individuals collected from Yunnan of southwestern China and Myanmar are grouped immediately next to the lineage of all domestic chickens, therefore *G. g. spadiceus* is considered as the closest wild progenitor to all modern chickens.



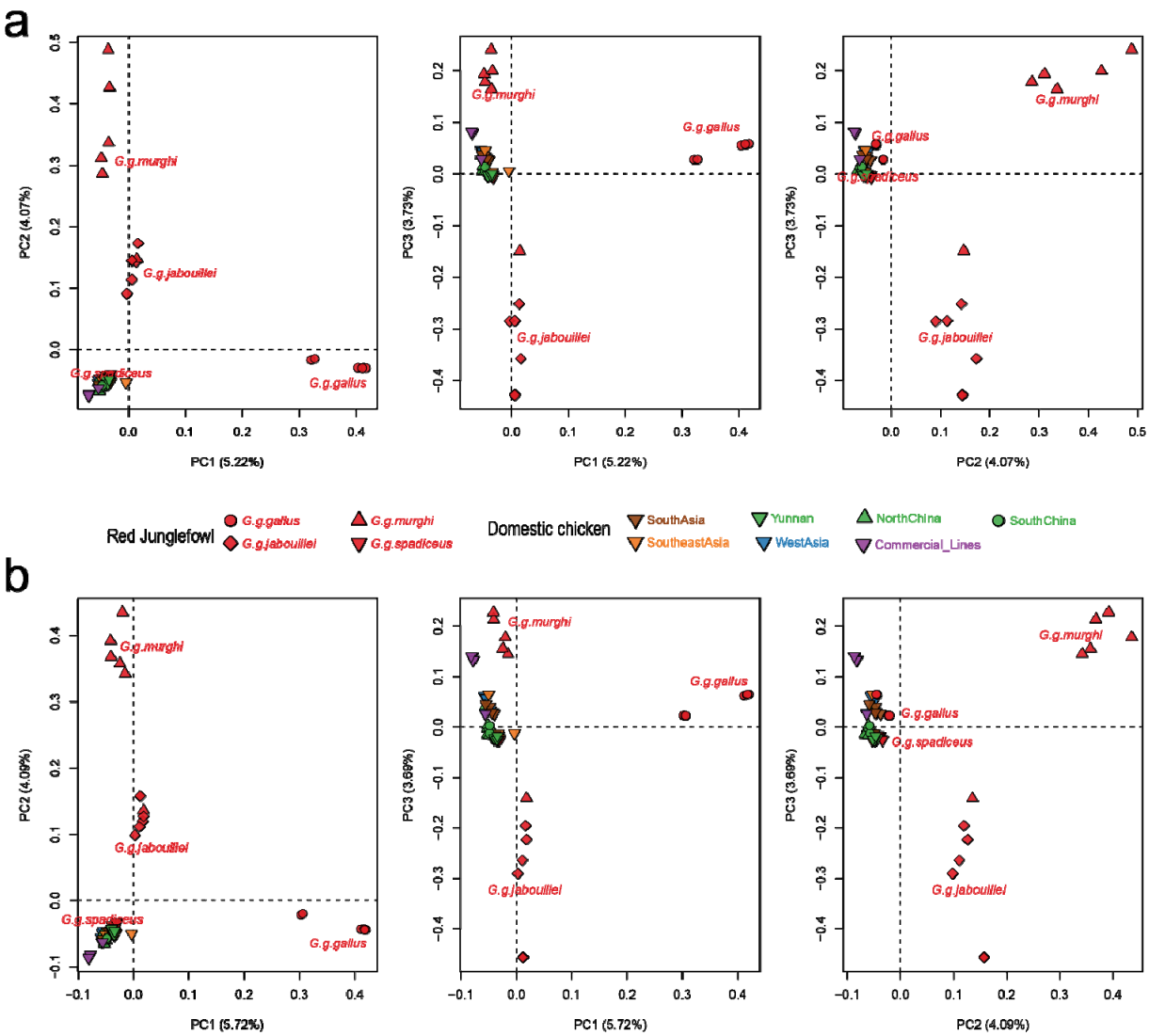
Supplementary information, Fig. S21.

Principal component (PC) analysis performed using smartpca program for all five Red Junglefowl subspecies of *G. g. bankiva*, *G. g. gallus*, *G. g. jabouillei*, *G. g. murghi* and *G. g. spadiceus* together with all domestic chickens. *G. g. spadiceus* is observed to be the most related to domestic chickens, whereas the other four Red Junglefowl subspecies are separated from all domestic chickens. The genetic cline is seen in the plot of PC2 against PC3 for *G. g. murghi* across its geographical range.



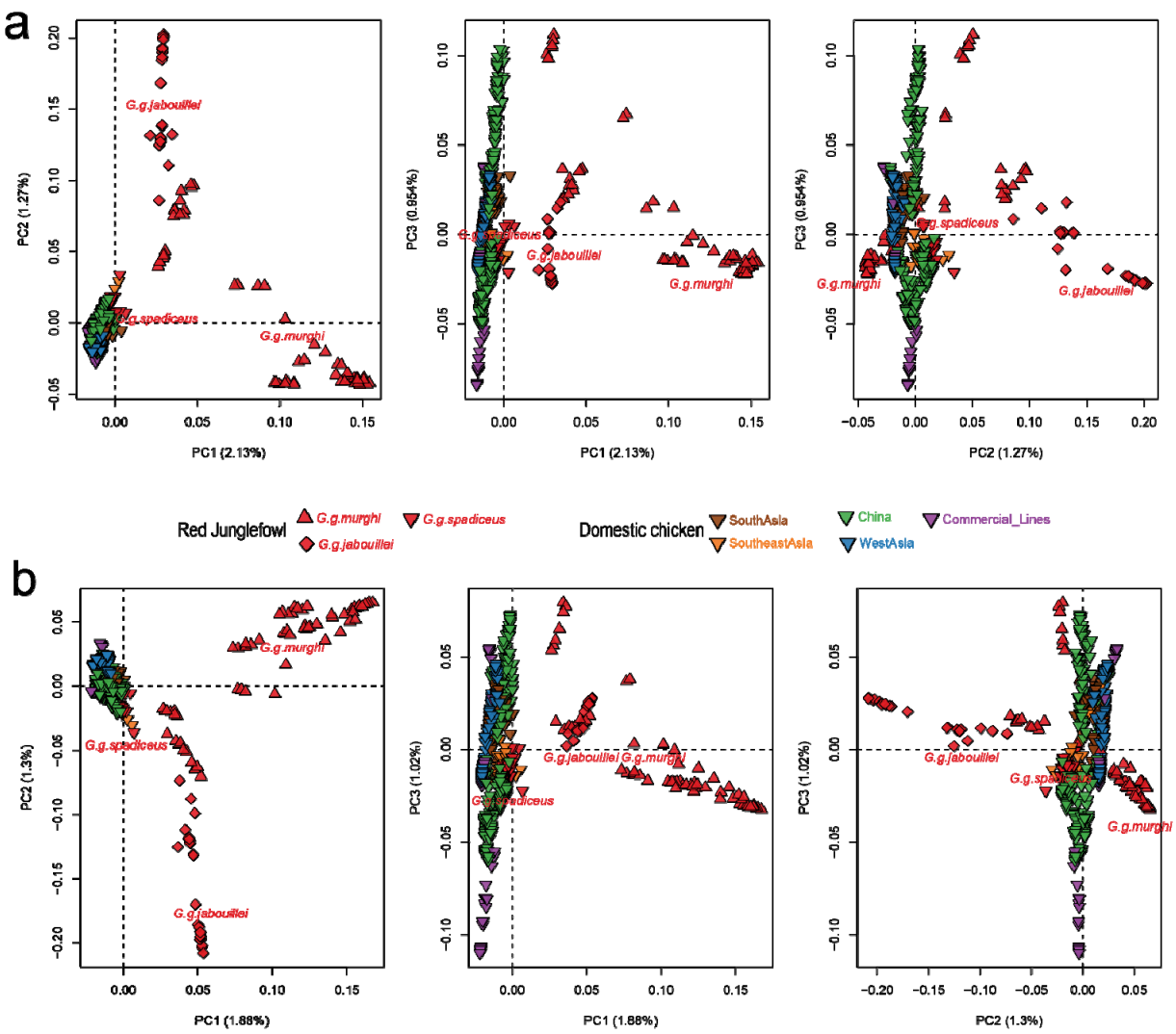
Supplementary information, Fig. S22.

Principal component (PC) analysis performed using GCTA software [16] for all five Red Junglefowl subspecies of *G. g. bankiva*, *G. g. gallus*, *G. g. jabouillei*, *G. g. murghi* and *G. g. spadiceus* together with all domestic chickens. The genetic cline is seen in the plot of PC2 against PC3 for *G. g. murghi* across its geographical range.



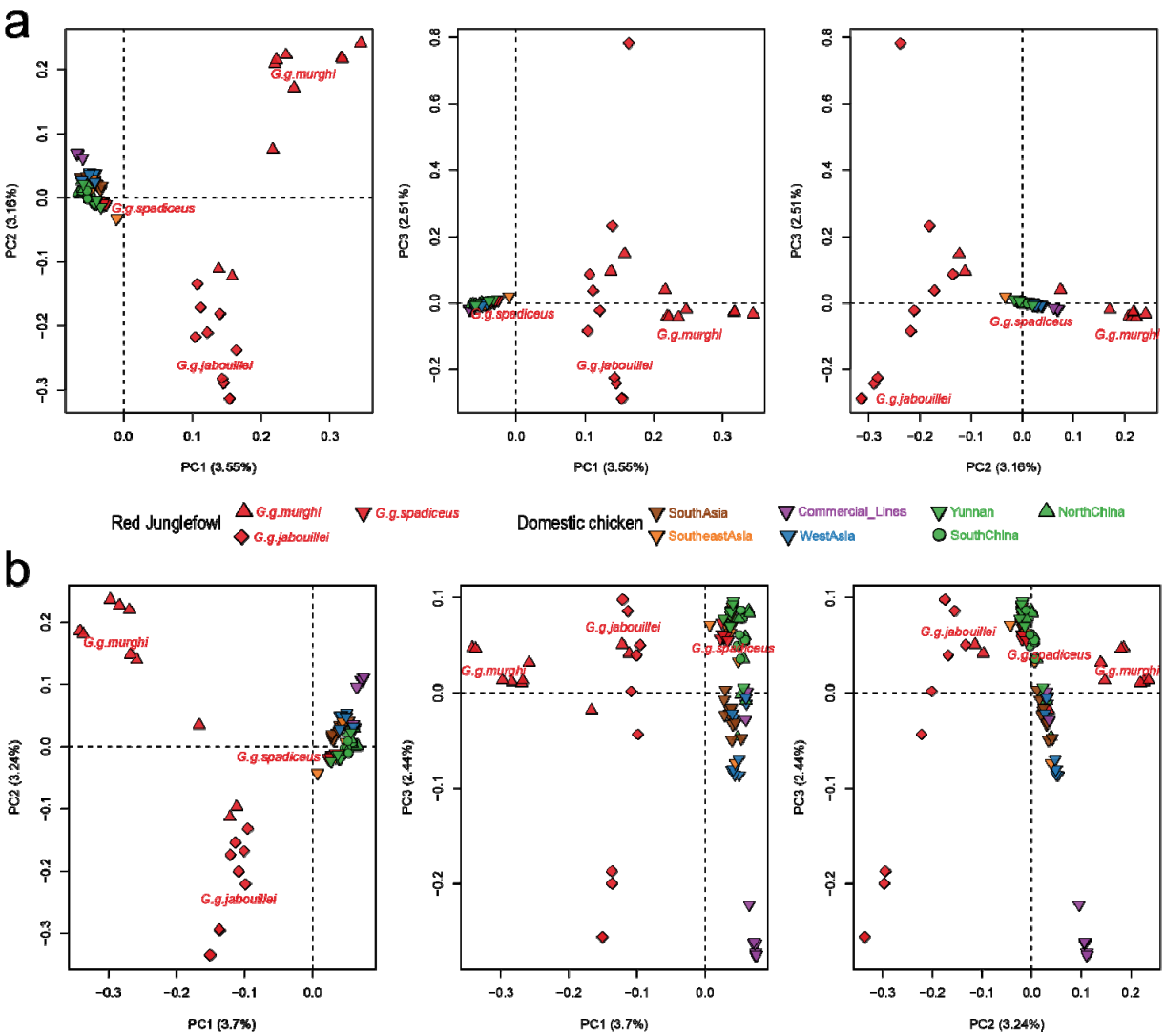
Supplementary information, Fig. S23.

Principal component (PC) analysis performed using smartpca program [17] for the four Red Junglefowl subspecies of *G. g. gallus*, *G. g. jabouillei*, *G. g. murghi* and *G. g. spadiceus* together with all domestic chickens. *G. g. bankiva* was removed from this analysis due to its high divergence from with other four Red Junglefowl subspecies. *G. g. spadiceus* is observed to be mixed with domestic chickens in all three plots. The genetic cline is seen in the plots of PC1 against PC2 and PC3 for *G. g. murghi* across its geographical range. Six randomly selected samples for each population were used for analysis.



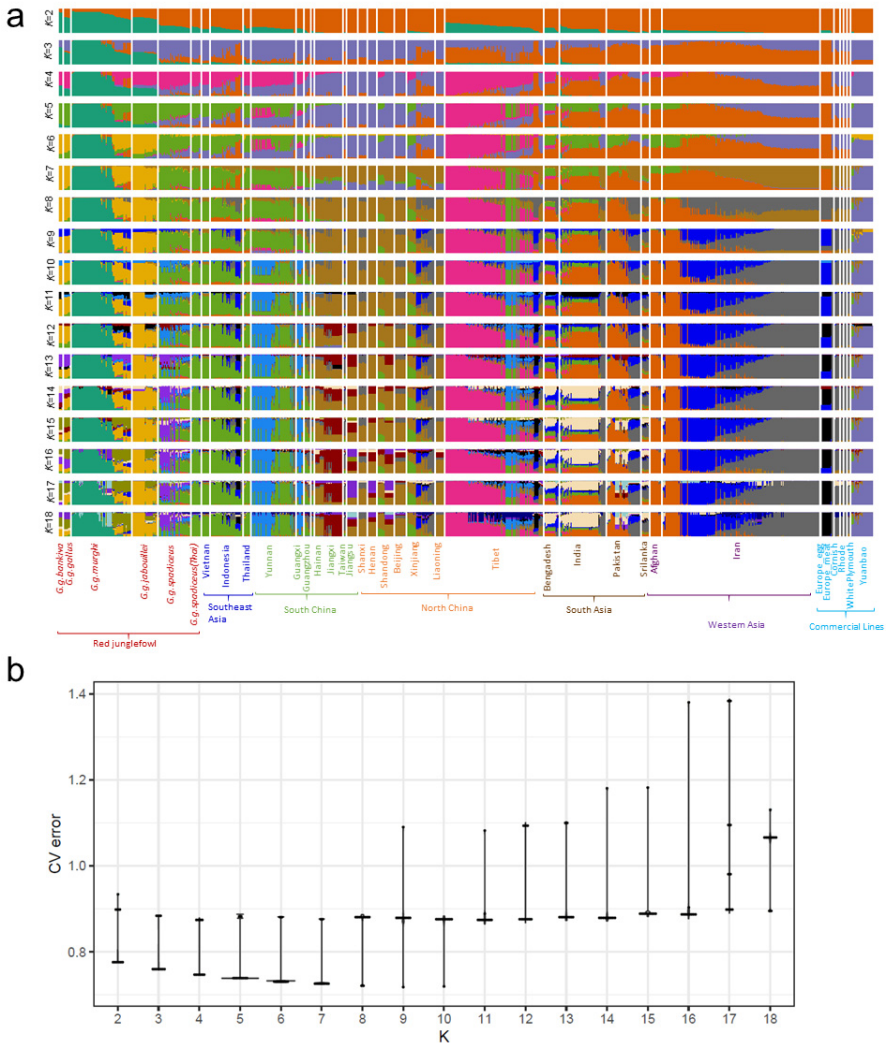
Supplementary information, Fig. S24.

Principal component (PC) analysis performed using GCTA software [16] for the four Red Junglefowl subspecies of *G. g. gallus*, *G. g. jabouillei*, *G. g. murghi* and *G. g. spadiceus* together with all domestic chickens. *G. g. bankiva* was removed from this analysis due to its high divergence from with other four Red Junglefowl subspecies. *G. g. spadiceus* is observed to be mixed with domestic chickens in all three plots. The genetic cline is seen in the plots of PC1 against PC2 and PC3 for *G. g. murghi* across its geographical range.



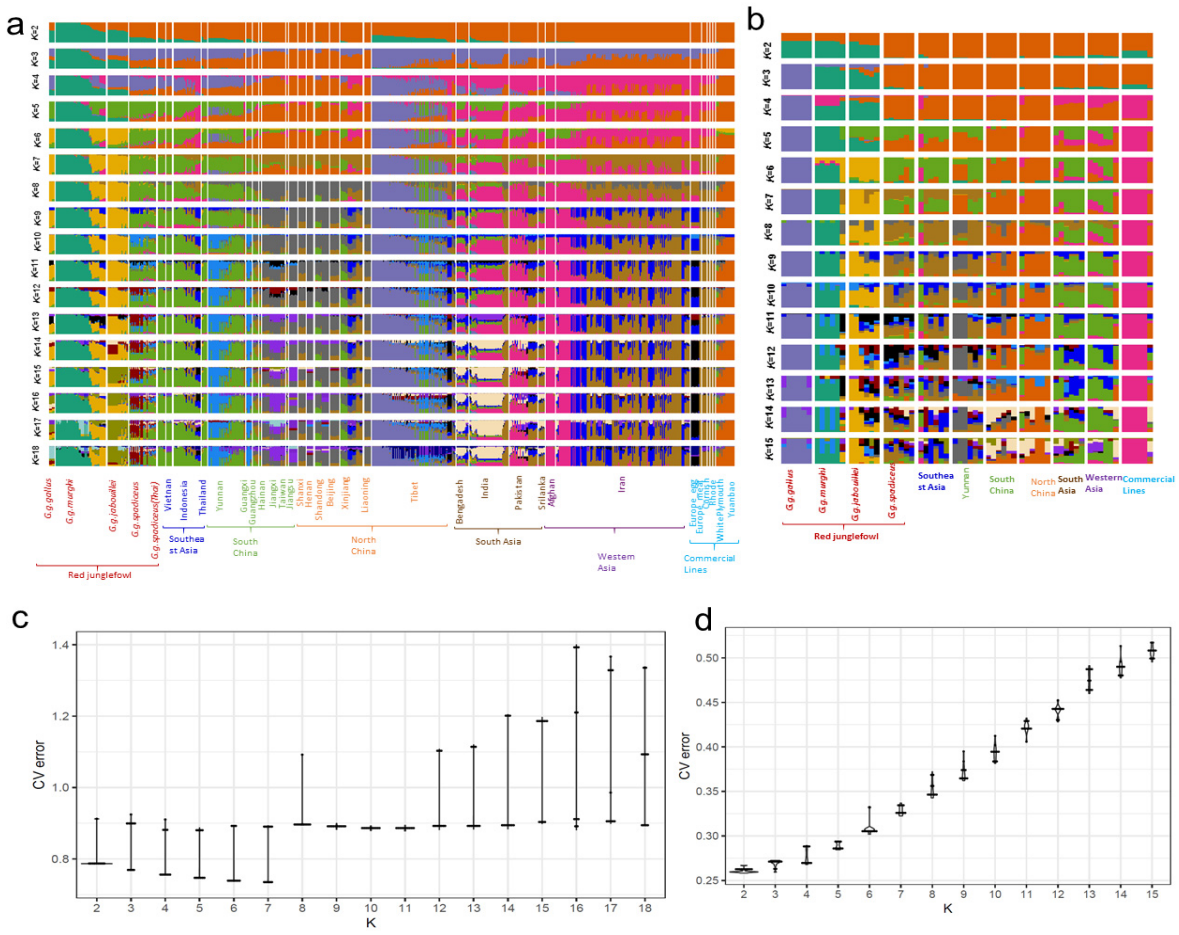
Supplementary information, Fig. S25.

Principal component (PC) analysis performed using smartpca program [17] for the three Red Junglefowl subspecies of *G. g. jabouillei*, *G. g. murghi* and *G. g. spadiceus* together with all domestic chickens. *G. g. bankiva* and *G. g. gallus* which are clearly separated from these three subspecies in all previous analyses were removed from this analysis to increase its resolution. Ten randomly selected samples for each population were used for analysis. *G. g. spadiceus* is observed to be mixed with domestic chickens in all three plots. The genetic cline is seen in the plots of PC1 against PC2 and PC3 for *G. g. murghi* across its geographical range.



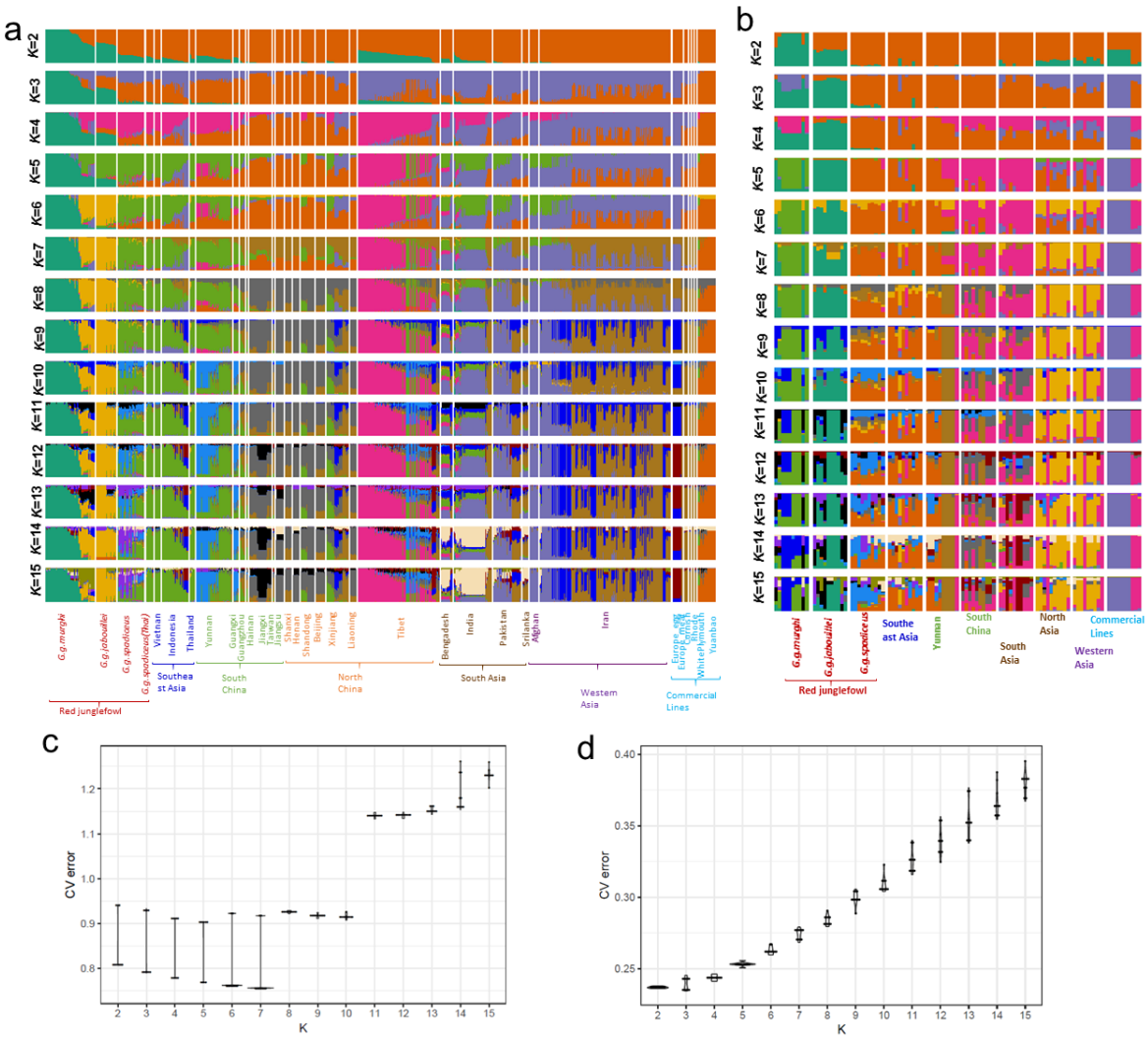
Supplementary information, Fig. S26.

The admixture analysis of genomes from all five Red Junglefowl subspecies (also a population of *G. g. spadiceus* collected from western Thailand, ThaiRJF) and all domestic chicken populations. Bar plots from top to bottom show *K* values from 2 to 20. The analysis is based on the autosomal variants pruned by PLINK toolset [77] with the following parameters "--indep 100 50 0.2". *G. g. bankiva* and *G. g. gallus* seem to share different amounts of genomic variations, though the patterns are not very stable across *K* values of and below 14, with other Red Junglefowl subspecies and some indigenous chicken populations. Their specific population genomic components for all Red Junglefowls and domestic chickens are established at $K = 15$ and beyond. The unstable admixture patterns observed for both *G. g. bankiva* and *G. g. gallus* is likely due to their limited samples included in this study.



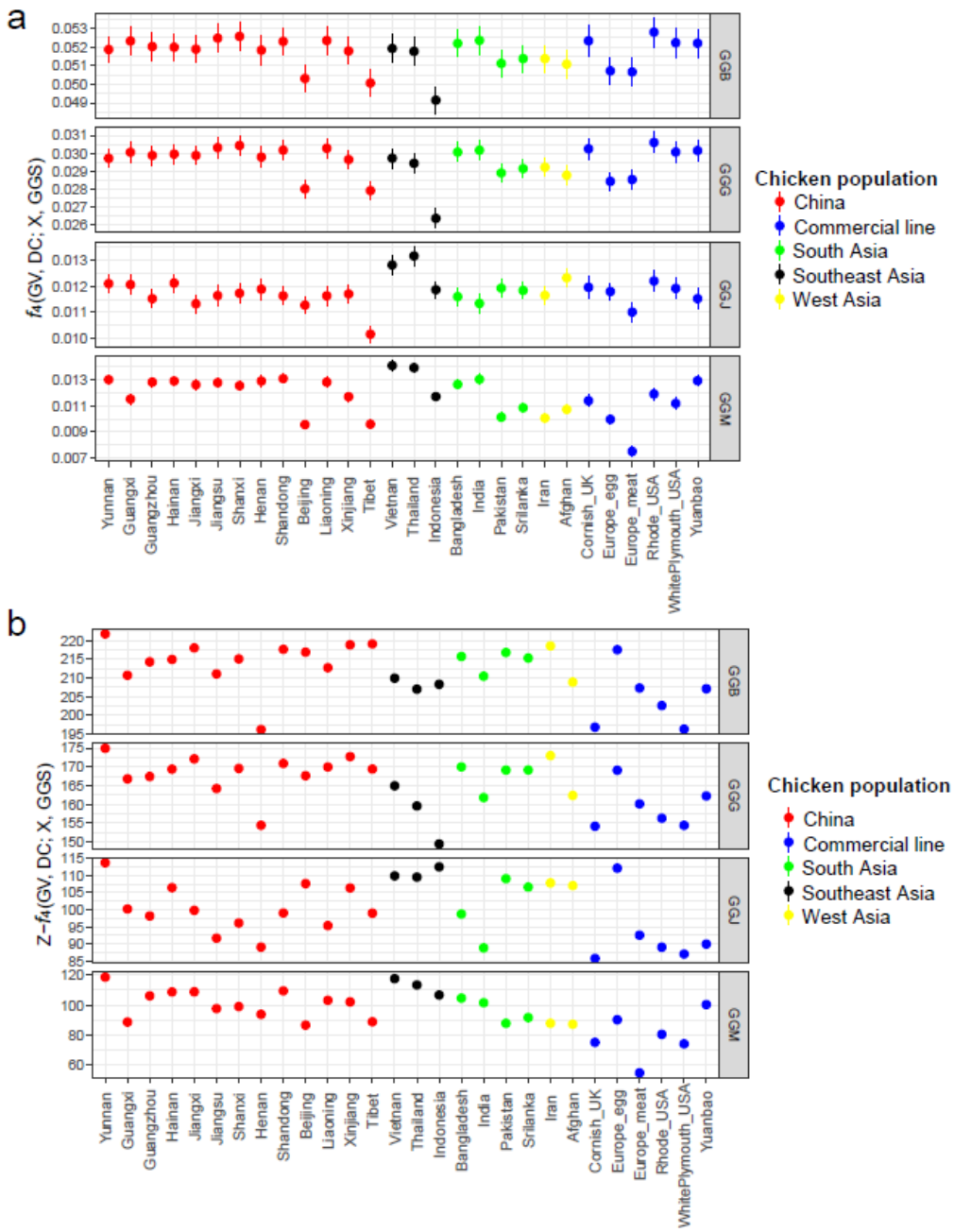
Supplementary information, Fig. S27.

The admixture analysis of genomes from four Red Junglefowl subspecies (also a population of *G. g. spadiceus* collected from western Thailand, ThaiRJF) and all domestic chicken populations. *G. g. bankiva* was removed from this analysis due to its high divergence from with other four Red Junglefowl subspecies. Bar plots from top to bottom denote *K* values ranging from 2 to 10. The analysis is based on the autosomal variants pruned by PLINK toolset [77] with the following parameters "--indep 100 50 0.2". **a, b** The analysis performed for all samples and six randomly selected samples for each subspecies, with CV error shown in **c** and **d**, respectively.

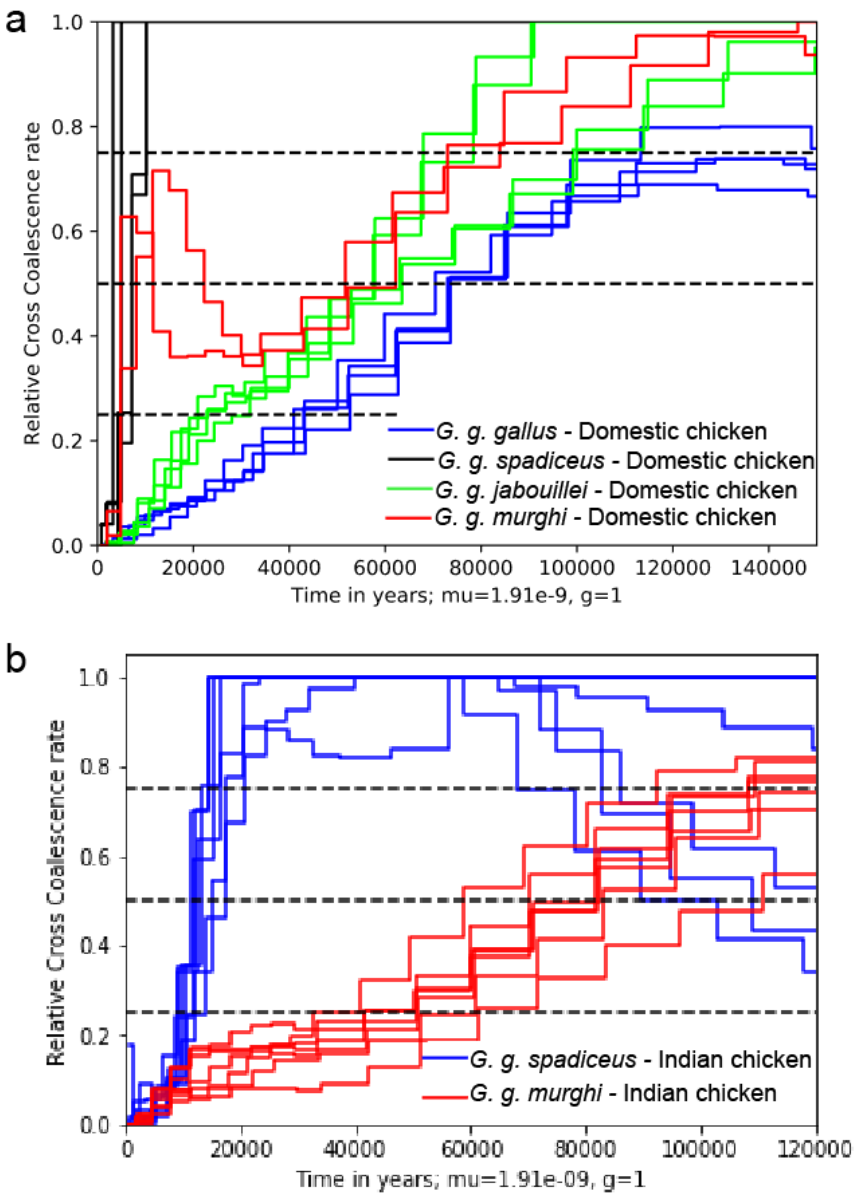


Supplementary information, Fig. S28.

The admixture analysis of genomes from three Red Junglefowl subspecies (also a population of *G. g. spadiceus* collected from western Thailand, ThaiRJF) and all domestic chickens. *G. g. bankiva* and *G. g. gallus* which are clearly separated from these three subspecies in all previous analyses were removed from this analysis to increase its resolution. Bar plots from top to bottom denote *K* values ranging from 2 to 10. The analysis is based on the autosomal variants pruned by PLINK toolset [77] with the following parameters "--indep 100 50 0.2". **a, b** The analysis performed for all samples and ten randomly selected samples for each subspecies, with CV error shown in **c** and **d**, respectively.

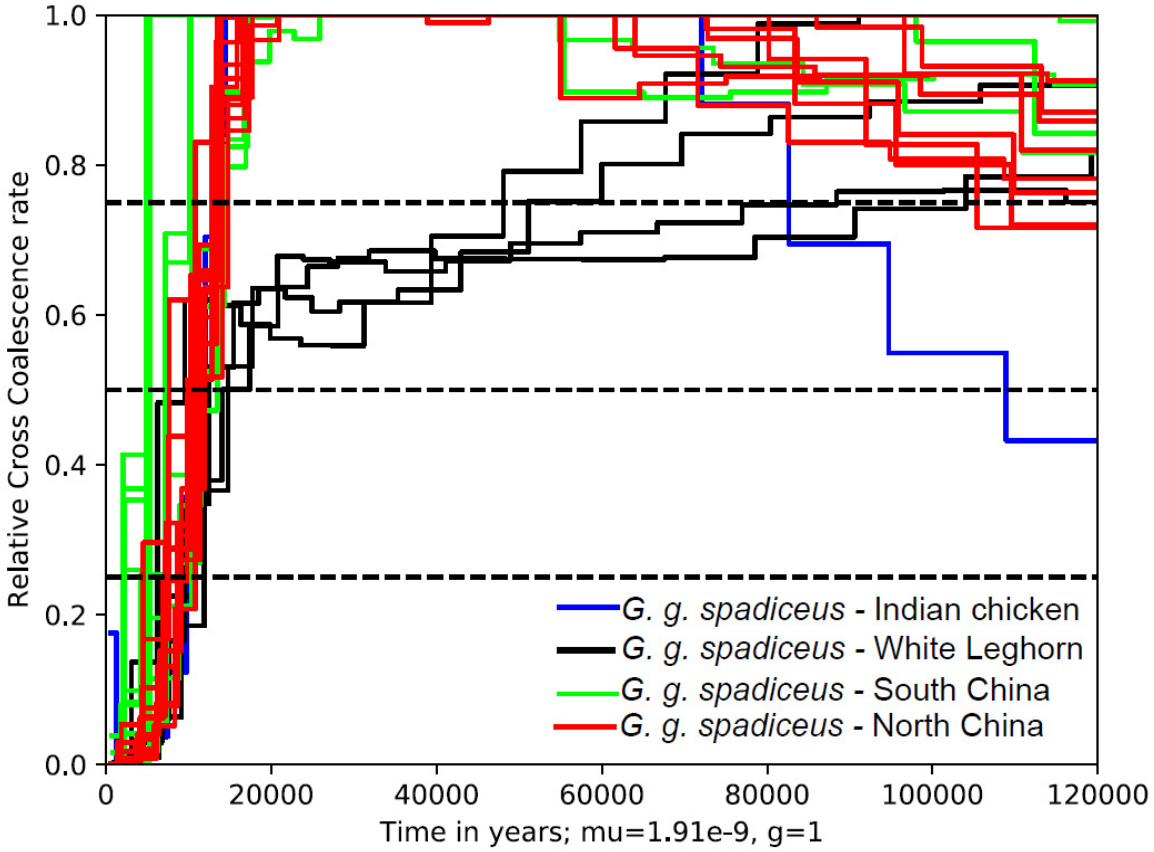


Supplementary information, Fig. S29.
*f*4 statistics in the form of *f*4(GV, DC; X, GGS) showing that domestic chickens share a closer genetic affinity and more predominant ancestry with *G. g. spadiceus* than with other RJF subspecies (*f*4 > 0; Z >> 3), where X represents RJF populations as shown on the top of each column and DC indicates domestic chickens grouped and colored according to their sampling locations. **a** *f*4 value. **b** Z score for each *f*4 calculated for each group. The estimated *f*4 value ± standard errors are plotted. GV, *G. varius*; GGS, *G. g. spadiceus*; GGM, *G. g. murghi*; GGG, *G. g. gallus*; GGJ, *G. g. jabouillei*; and GGB, *G. g. bankiva*.



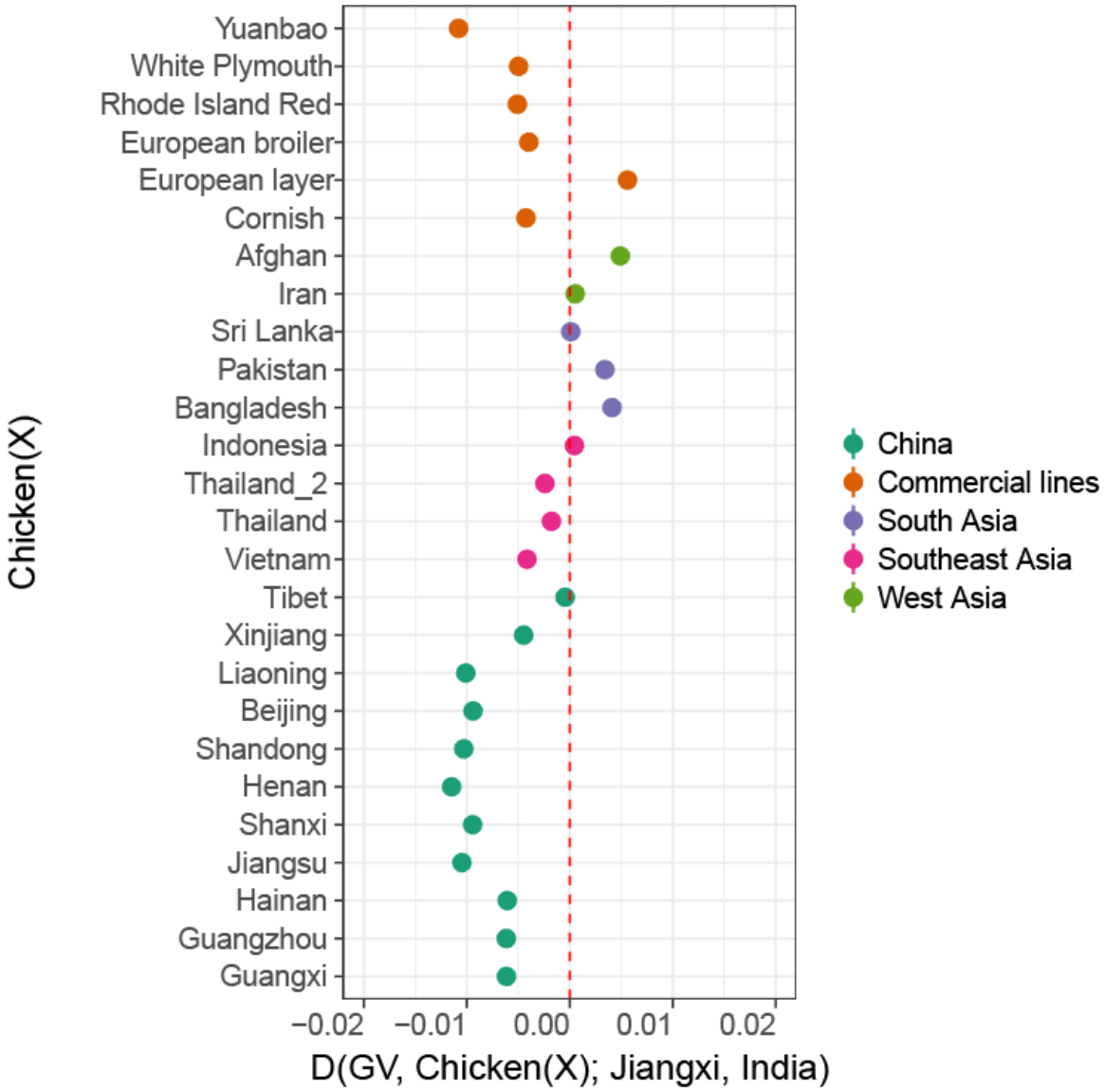
Supplementary information, Fig. S30.

The MSMC analysis reveals the splitting of domestic chickens from Red Junglefowl subspecies. **a** The divergence of domestic chickens at the base of phylogenetic trees (Fig. 2a) from each Red Junglefowl subspecies as measured by CCR at 50%. **b** The divergence of Indian chickens from either *G. g. murghi* or *G. g. spadiceus.* It is clear that Indian chickens diverged from *G. g. murghi* much earlier than the domestication of *G. g. spadiceus*.



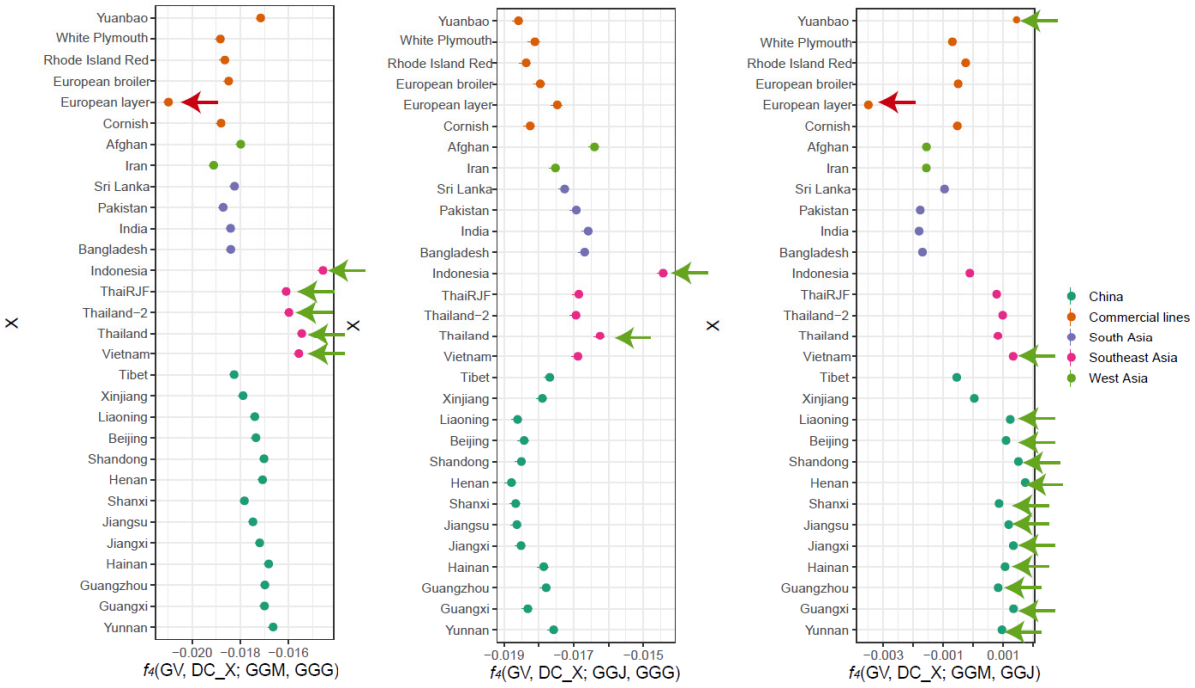
Supplementary information, Fig. S31.

The MSMC analysis demonstrates the divergence of domestic chickens from *G. g. spadiceus* around 9,500 ± 3,300 years ago.



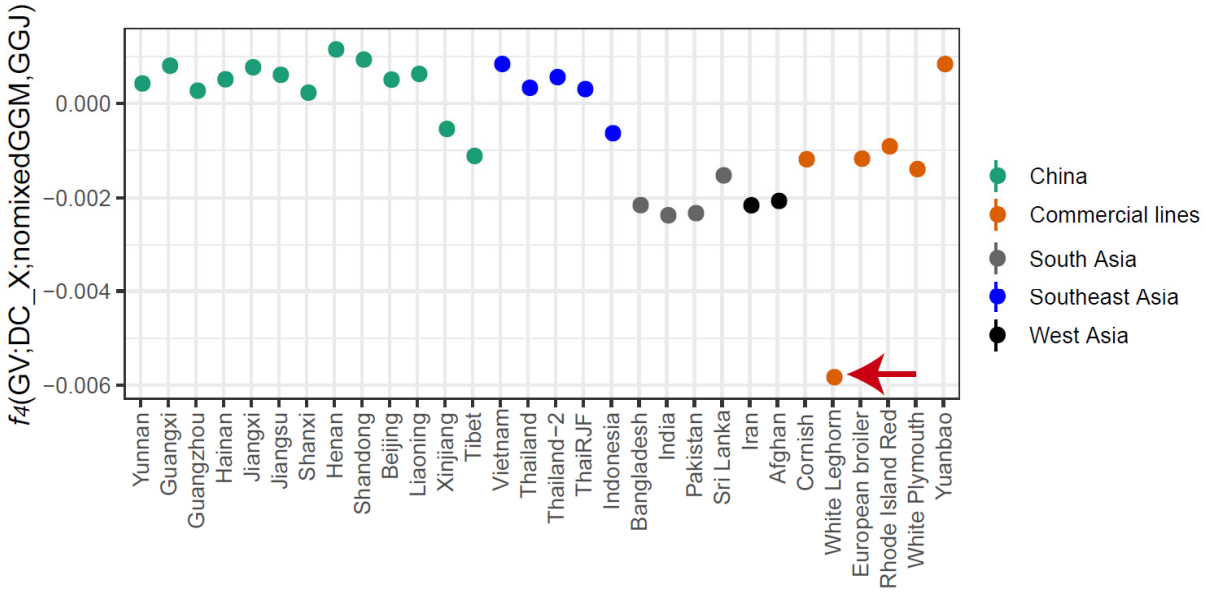
Supplementary information, Fig. S32.

D-statistics in form of *D*(GV, Chicken_X; Core1, Core2) indicating the average genetic distance for each chicken population (Chicken_X) to "two core" chicken populations. Core1 is the chickens from Jiangxi province in southern China (clade II in Fig. 2a) and Core2 is Indian chickens (clade I in Fig. 2a). Positive *D* values suggest a closer relationship for chicken population X with Indian chickens while a negative value indicates a closer relationship with southern Chinese chickens. GV, Green Junglefowl.



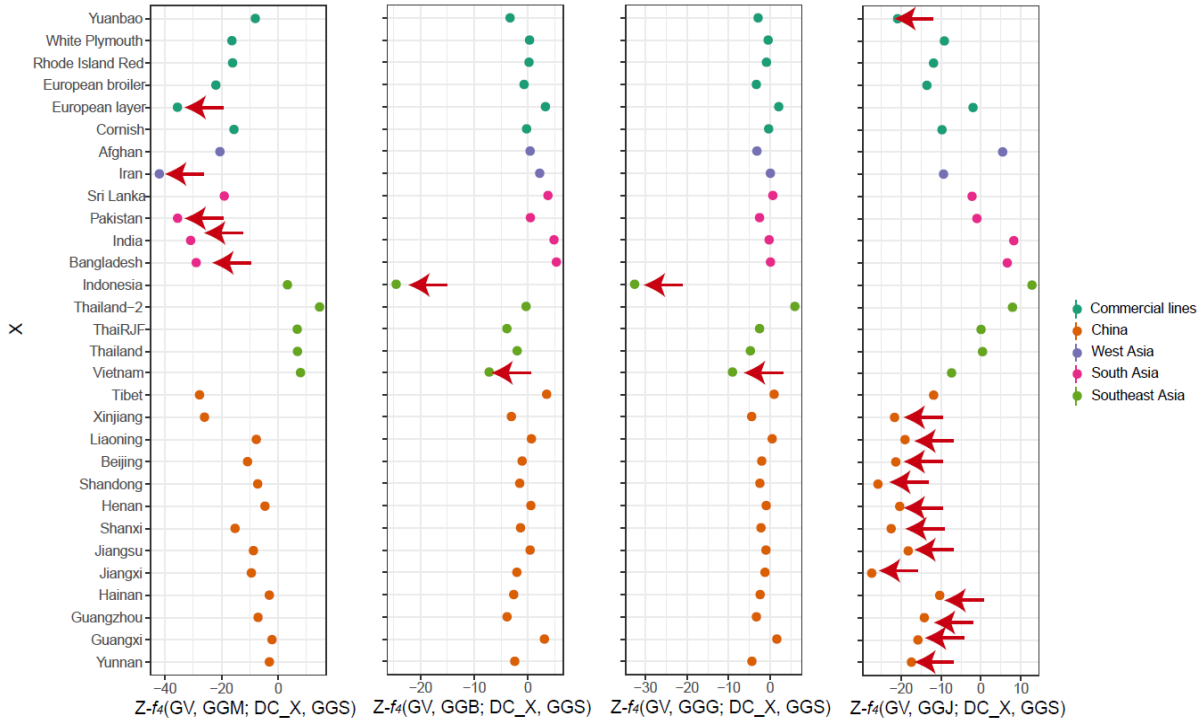
Supplementary information, Fig. S33.

The f_4 statistics illustrate that some indigenous chicken populations distributed close to the geographic range of a particular Red Junglefowl subspecies carry a higher proportion of their ancestry from such subspecies, providing evidence of occasional and localized hybridizations between indigenous chickens and wild Red Junglefowls. The significant signature of admixture is marked with arrows. For example, *G. g. murghi* (GGM) shows the signature of admixture with Indian, Pakistani, Iranian and Tibetan chickens, as well as with White Leghorn breed (European layer); *G. g. jabouillei* (GGJ) displays signature of admixture with Chinese chickens; *G. g. gallus* (GGG) carries signatures of admixture with Indonesian and Thai chickens. Populations with significant signatures of admixture are marked using arrows. GV, *G. varius*; DC_X, domestic chickens grouped according to their sampling locations and also commercial lines as shown in vertical axis.



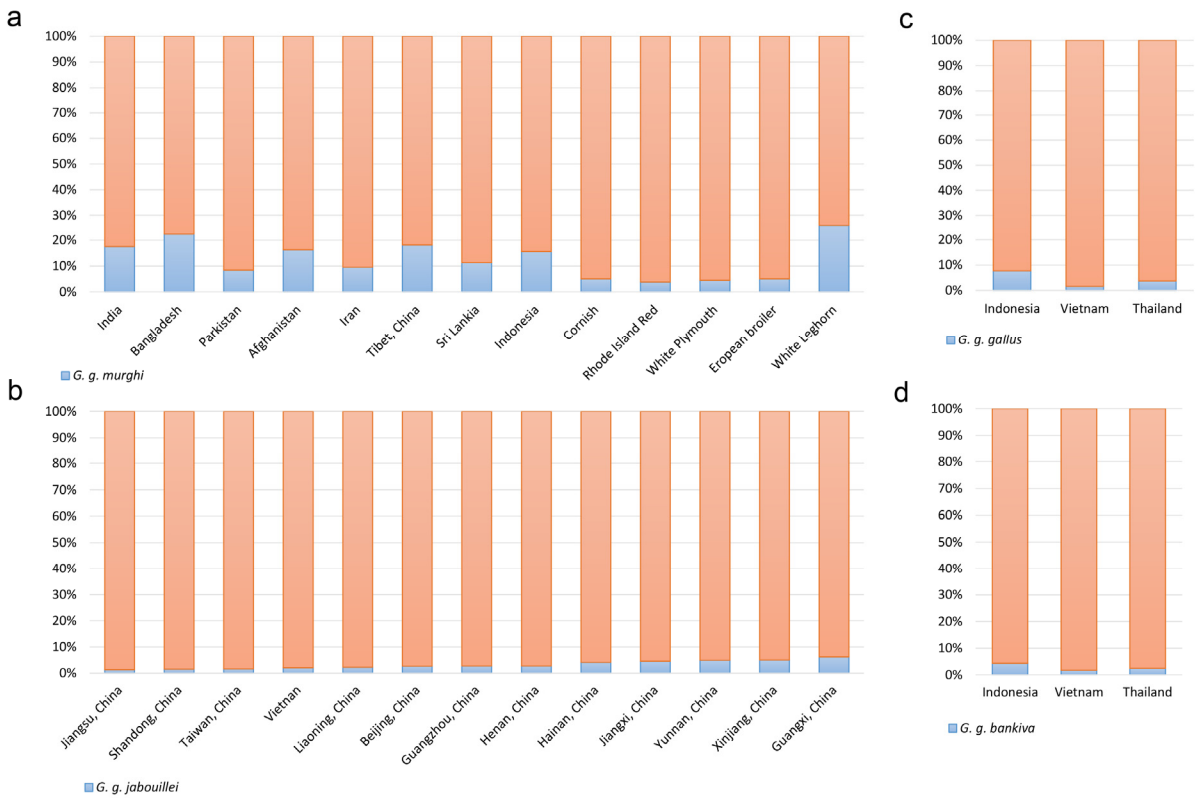
Supplementary information, Fig. S34.

The f_4 statistics illustrate that some indigenous chicken populations distributed close to the geographic range of Indian and South Chinese wild Red Junglefowl subspecies carry a higher proportion of their ancestry from such subspecies, supporting occasional and localized hybridizations between indigenous chickens and wild Red Junglefowls (*G. g. murghi* (GGM) in Indian and *G. g. jabouillei* (GGJ) in Southern China). For GGM included in this analysis, we removed birds with admixed genomes according to the PCA and admixture results shown in Supplementary information, Figs. S21-S28. It is clear that GGM shows signature of admixture with Indian, Pakistani, Iranian and Tibetan chickens while GGJ has signatures of admixture with Chinese chickens. It is interesting to note that the f_4 statistic for White Leghorn becomes more significant after removing a few putatively admixed GGM, a further confirmation of the admixture between *G. g. murghi* and White Leghorn breed. GV, *G. varius*; DC_X, domestic chickens grouped according to their sampling locations and commercial lines as shown in horizontal axis.



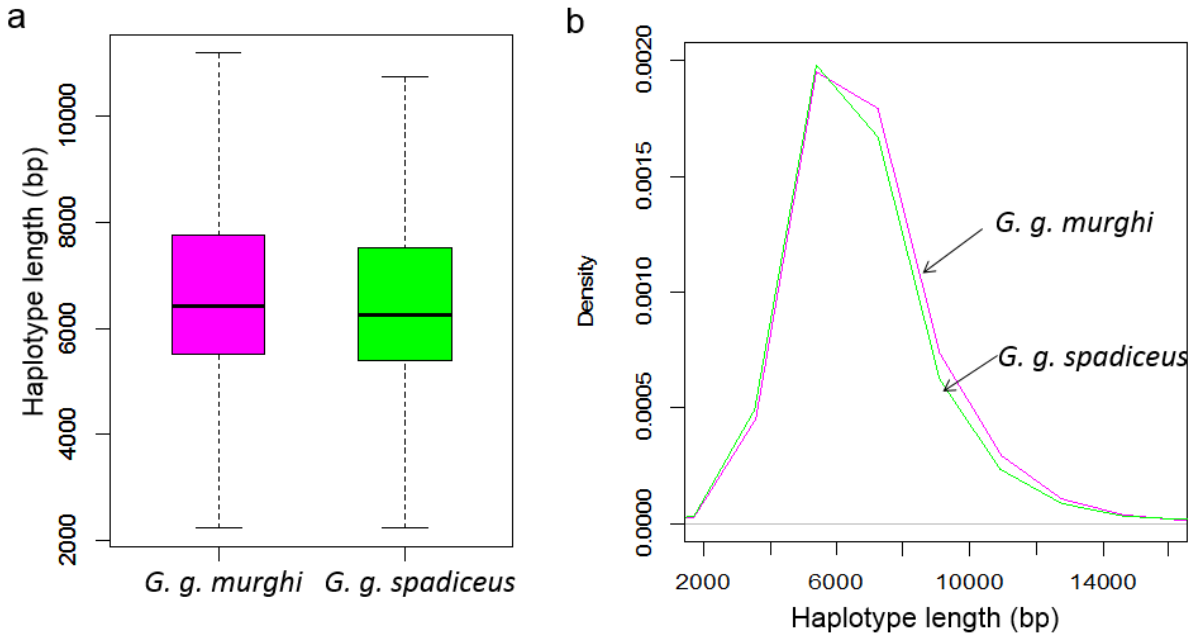
Supplementary information, Fig. S35.

The f_4 statistics illustrate localized hybridizations between Red Junglefowls and domestic chickens. The European layer used here is of White Leghorn breed. The significant signatures of admixture are marked with arrows. For example, *G. g. murghi* (GGM) show signatures of admixture with Indian, Pakistani, Iranian and Tibetan chickens; *G. g. jabouillei* (GGJ) with Chinese chickens; *G. g. bankiva* (GGB) with Indonesian chickens; and *G. g. gallus* (GGG) with Southeast Asian chickens. GV, *G. varius*; DC_X, domestic chickens grouped according to their sampling locations and commercial lines as shown in vertical axis.



Supplementary information, Fig. S36.

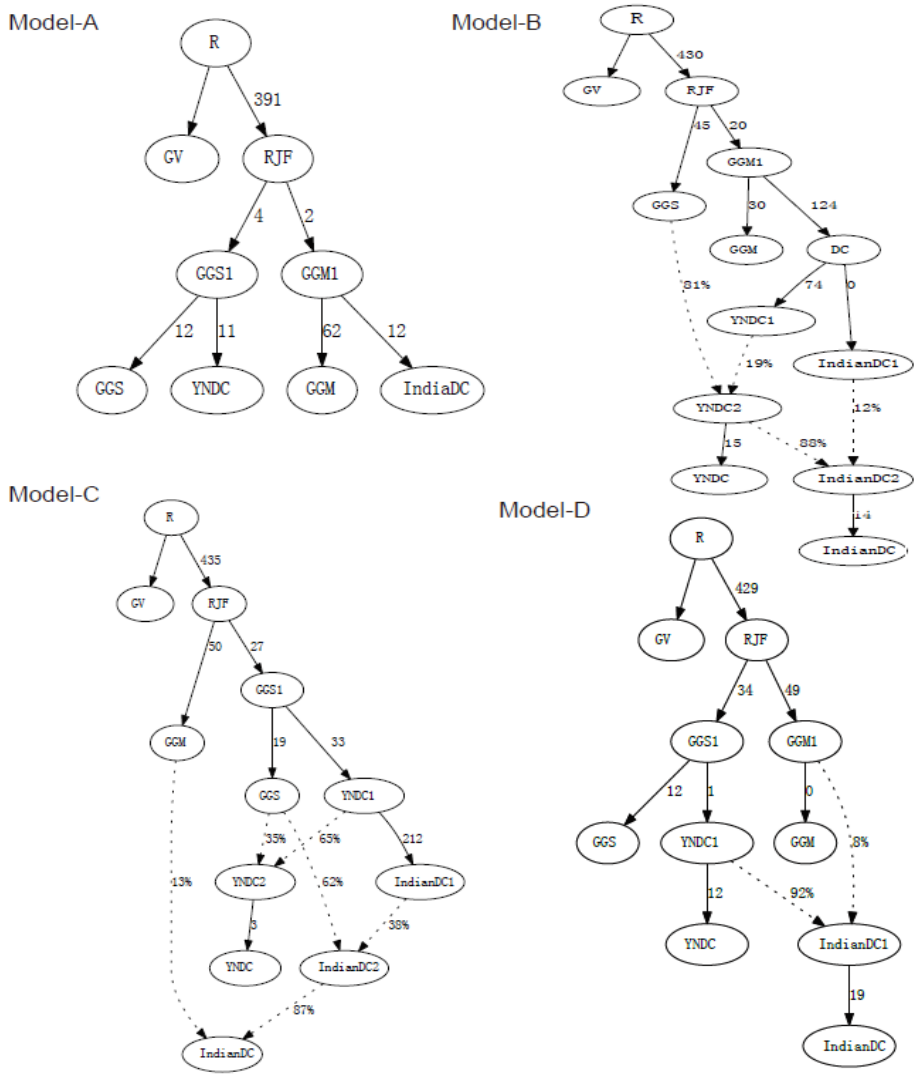
The proportion of genome regions introgressed between domestic chickens and Red Junglefowl subspecies estimated by f_4-ratio. **a-d** Specific amounts of such gene flow with *G. g. murghi*, *G. g. jabouillei*, *G. g. gallus* and *G. g. bankiva*, respectively.



Supplementary information, Fig. S37.

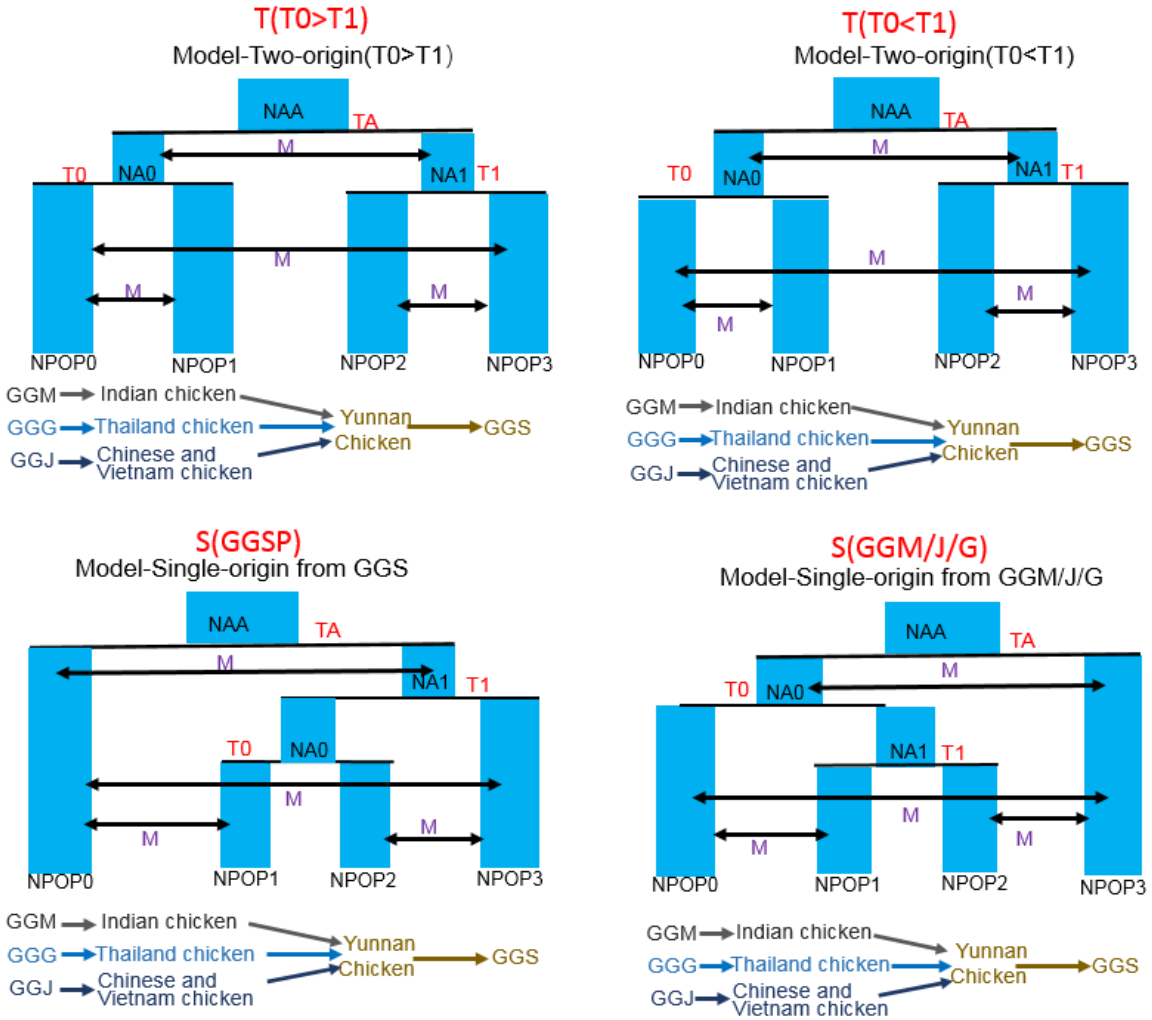
Indian chickens retain longer admixed genomic segments from *G. g. murghi* than those from *G. g. spadiceus*, implying a more recent introgression event from *G. g. murghi* into South Asian chickens. **a** Haplotype length distribution in box plots of the introgressed segments from *G. g. murghi* or *G. g. spadiceus* into Indian chickens. **b** Plot of the haplotype length against the density of the introgressed segments from *G. g. murghi* or *G. g. spadiceus* into Indian chickens.

PCAdmix program (version: 1.0) [32] was used to perform this local ancestry inference. *G. g. murghi*, *G. g. spadiceus* and Yunnan chickens were used as source populations while Indian chickens were treated as an 'admixed population'.

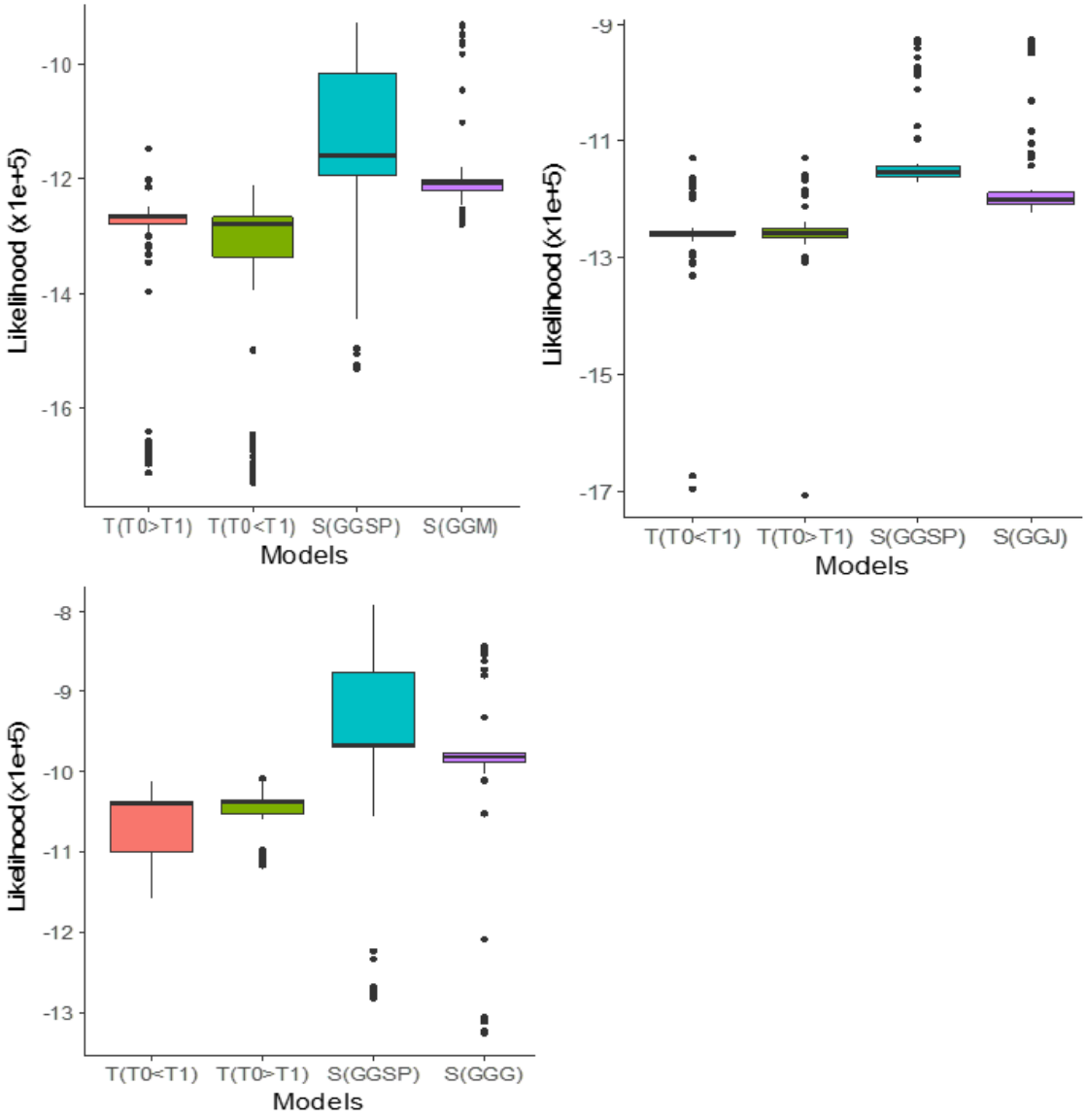


Supplementary information, Fig. S38.

Scaffold admixture models used to determine the origin of domestic chickens in India (IndianDC) and/or Yunnan in southwestern China (YNDC; clustered in the basal position of other domestic chickens as shown in the phylogenetic tree). Dotted lines denote admixture events and values besides the dotted lines correspond to admixture proportions. Branch lengths are shown in units of F_{ST} × 1000. GV, *G. varius*; GGS, *G. g. spadiceus*; GGM, *G. g. murghi*; DC, domestic chicken. Model-A, YNDC, and IndianDC originated from *G. g. spadiceus* and *G. g. murghi*, respectively; Model-B, IndianDC derived first from *G. g. murghi* in India followed by their dispersal and then continuous admixture with *G. g. spadiceus* in Yunnan of China; Model-C, YNDC originated first from *G. g. spadiceus* in Yunnan of China followed by their dispersal and then continuous admixture from *G. g. murghi* in India; Model-D, IndianDC originated from the dispersed YNDC derived first from *G. g. spadiceus* in Yunnan of China followed by their relatively recent admixture with *G. g. murghi* several generations ago in India. Only model-C best fits the data, having no f_4 outliers, unlike Model-A, Model-B and Model-D. All empirical *f*-statistics greater than 3 standard errors are defined as outliers.

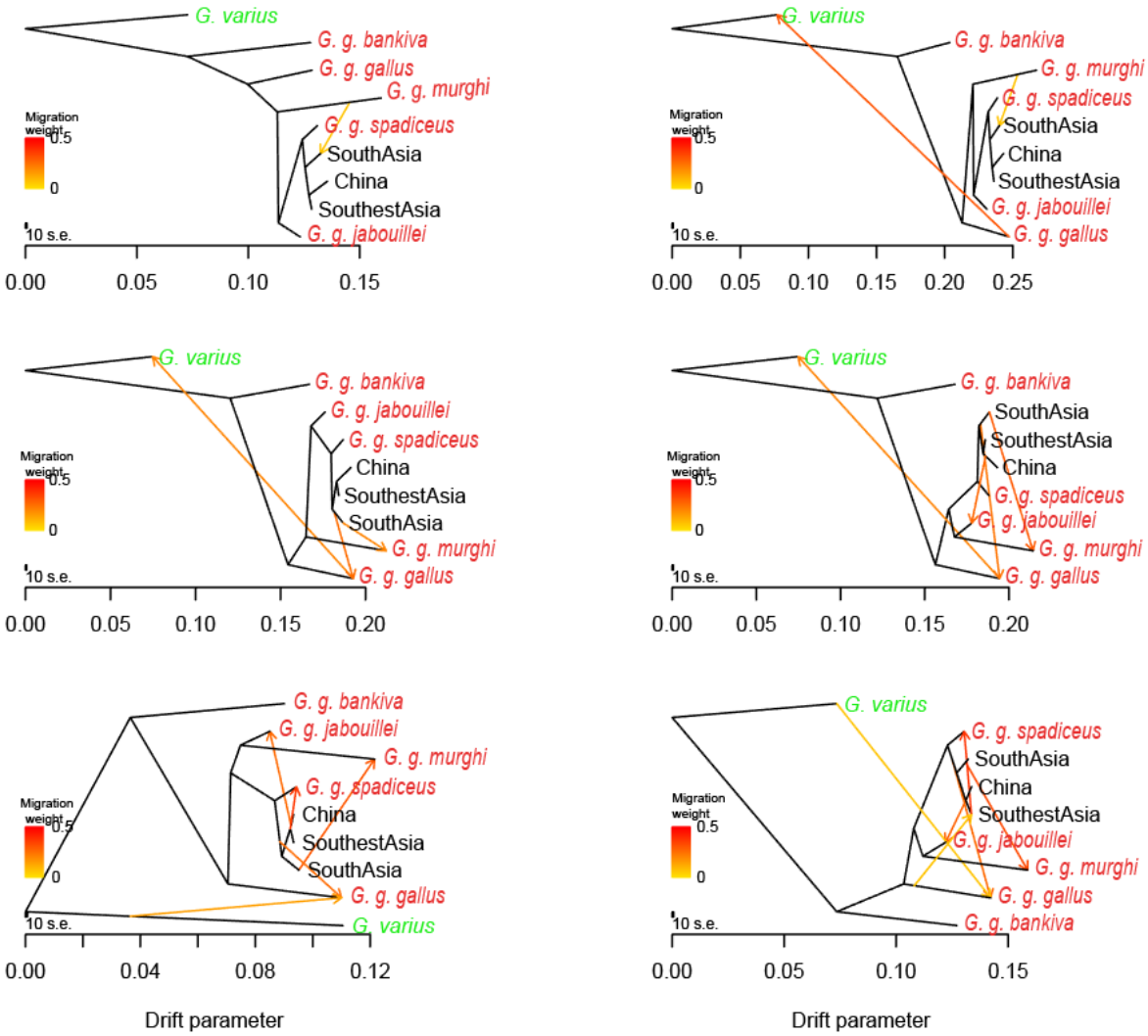


Supplementary information, Fig. S39.
Schematic representation of the proposed models regarding multiple origins and single origin with lateral introgression of domestic chickens tested using fastsimcoal2 software. Model-T (top) proposes a dual-origin of chicken domestication from *G. g. spadiceus* (GGS) at T1 together with *G. g. murghi* (GGM), *G. g. jabouillei* (GGJ) or *G. g. gallus* (GGG) at T0. Model-S (GGSP) (bottom left) considers a single origin of chicken domestication from *G. g. spadiceus* followed by migration and interbreeding with either *G. g. murghi* or *G. g. gallus* during their dispersal. Model-S (GGM/J/G) (bottom right) assumes a single origin of chicken domestication from *G. g. murghi*, *G. g. jabouillei* or *G. g. gallus* followed by interbreeding with *G. g. spadiceus* during their dispersal. T parameters correspond to the splitting time of populations. N and NA parameters indicate current and ancestral effective population sizes, respectively. Arrows indicate migrations between two populations, as noted by 'M'.



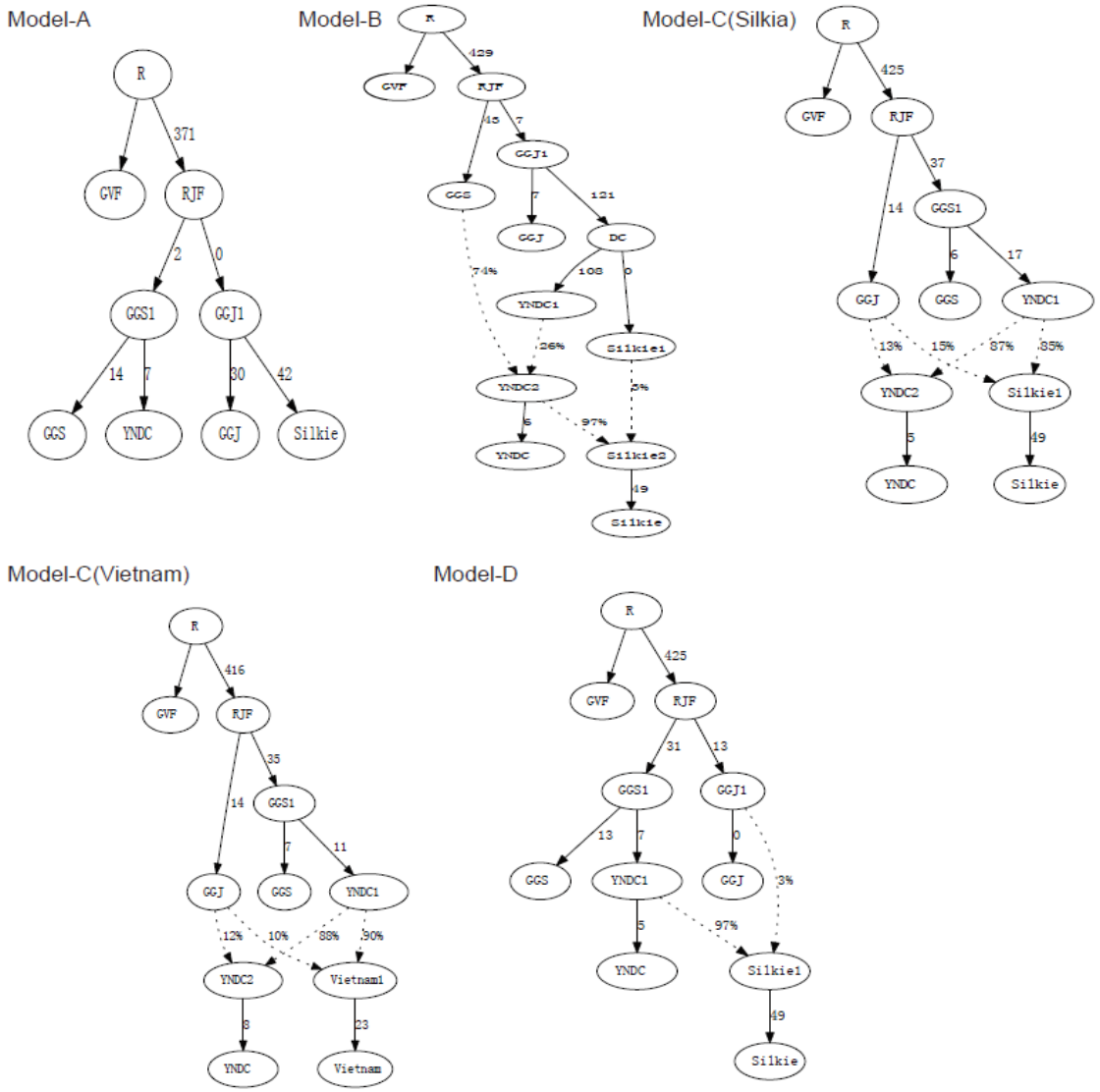
Supplementary information, Fig. S40.

Comparison of likelihood estimates generated under different models. T(T0<T1) and T(T0>T1) models indicate that domestic chicken originated twice with T0<T1 and T0>T1 as depicted in Supplementary information, Fig. S39, respectively. S(GGSP), S(GGM), S(GGJ), and S(GGG) assume single origin of domestic chicken from *G. g. spadiceus*, *G. g. murghi*, *G. g. jabouillei* and *G. g. gallus*, respectively, in the models. By comparing with other models, the model S(GGSP) has the highest likelihood across the three comparisons, supporting a single origin of domestic chicken from *G. g. spadiceus*, rather than from *G. g. murghi*, *G. g. jabouillei* or *G. g. gallus*.



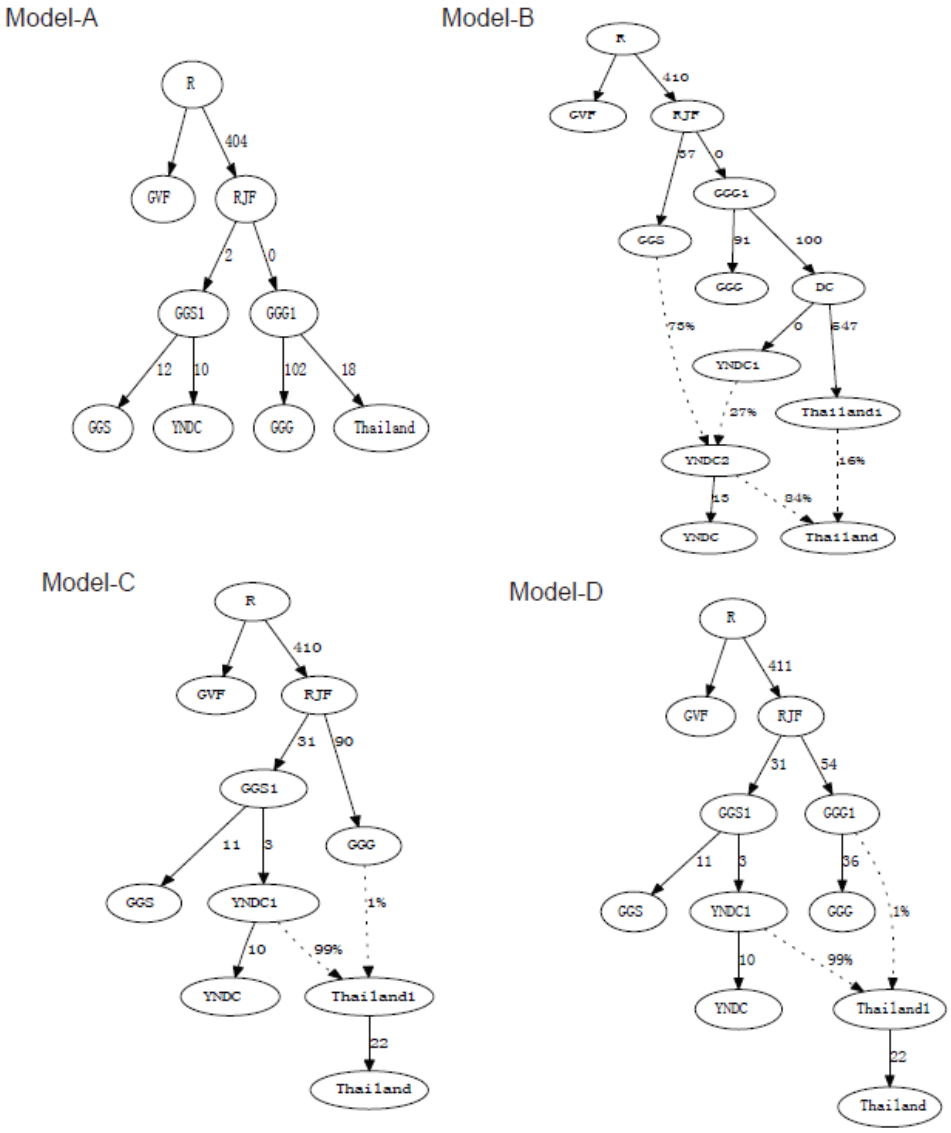
Supplementary information, Fig. S41.

TreeMix analysis (migration edges, M, ranging from 1 to 6) revealing the separation and admixture of Red Junglefowl and domestic chicken populations. M = 1 and 2 explain 98.5% and 99.7% of total variation, respectively. By increasing M, the fit for the model improved marginally. Population names marked in back and red denote domestic chicken and Red Junglefowls, respectively. The Green Junglefowl (*G. varius*) was used as an outgroup in this analysis. Indigenous chickens are grouped into three big groups: South Asia (Bangladesh and India), China (Peking, Shandong, Liaoning, Shanxi, Jiangsu and Jiangxi provinces) and Southeast Asia (Yunnan and Guangxi provinces of China, Thailand and Vietnam). All domestic chickens form a monophyletic group across six migration edges, with further diversification into two branches, one presenting SouthAsian chickens and another containing chickens from Southeast Asia and China. We observed an early split of *G. g. gallus, G. g. murghi* and *G. g. jabouillei* from domestic chickens while all domestic chickens are descendants of the same source population that split off from *G. g. spadiceus*.



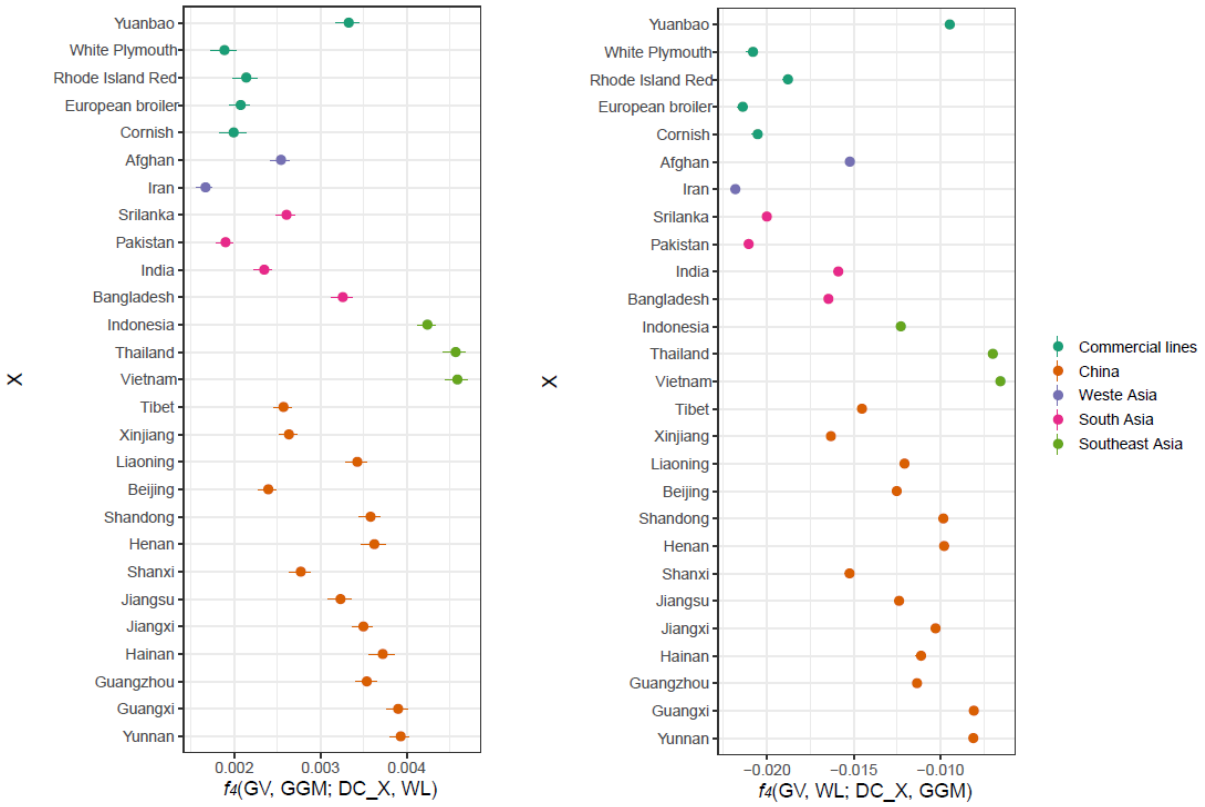
Supplementary information, Fig. S42.

Scaffold admixture graph used for modeling the origin of domestic chickens in South China and/or Yunnan in southwestern China. Dotted lines denote admixture events and values besides the dotted line correspond to admixture proportions. Branch lengths are shown in units of $F_{ST} \times 1000$. GV, *G. varius*; GGS, *G. g. spadiceus*; Silkie, a famous chicken breed from southern China; DC, domestic chicken; and GGJ, *G. g. jabouillei*. YNDC, local chicken in Yunnan basal to other domestic chickens as shown in the phylogenetic tree. Model-A, chickens originated from *G. g. spadiceus* and *G. g. jabouillei* independently. Model-B, all chickens originated only from *G. g. jabouillei* while the primary chickens (YNDC1) in Yunnan experienced a continuous admixture with *G. g. spadiceus* in southwestern China. Model-C (Silkie), all chickens originated only from *G. g. spadiceus* while chickens in southern China were developed from the primary chickens dispersed from Yunnan through interbreeding with *G. g. jabouillei*. Model-C (Vietnam), all chickens originated only from *G. g. spadiceus* while chickens in Vietnam were developed from the primary chickens dispersed from Yunnan through interbreeding with *G. g. jabouillei* over time. Model-D, chickens in southern China were developed from the primary chickens dispersed from Yunnan through interbreeding with *G. g. jabouillei* several generations ago. Model-C for chickens in southern China and Vietnam fit the data best with no f_4 outliers, unlike Model-A, Model-B, and Model-D. All empirical *f*-statistics greater than 3 standard errors are defined as outliers.



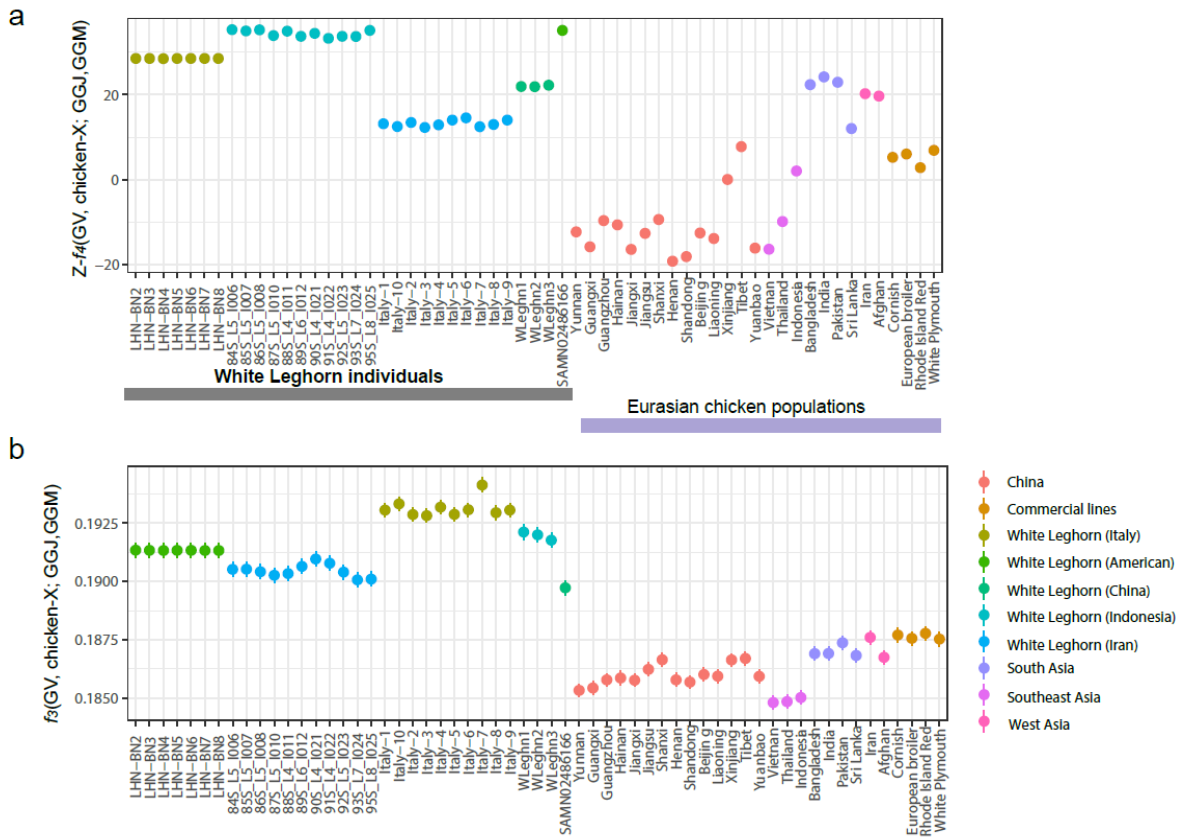
Supplementary information, Fig. S43.

Scaffold admixture models showing domestic chickens through single or multiple events in southern Thailand and neighboring islands (from *G. g. gallus*) as well as in Yunnan of southwestern China (from *G. g. spadiceus*). Dotted lines denote admixture events and values besides the dotted lines correspond to admixture proportions. Branch lengths are shown in units of $F_{ST} \times 1000$. GV, *G. varius*; GGS, *G. g. spadiceus*; GGG, *G. g. gallus*; DC, domestic chicken; YNDC, local chicken from Yunnan (also clustered in basal position of other domestic chickens as shown in phylogenetic tree). Model-A, all chickens originated from *G. g. spadiceus* and *G. g. gallus* independently; Model-B, all chickens originated only from *G. g. gallus* followed by interbreeding with *G. g. spadiceus* after their dispersal in Yunnan; Model-C, all chickens originated only from *G. g. spadiceus* followed by interbreeding with *G. g. gallus* after their migration to Southeast Asia; Model-D, all chickens originated only from *G. g. spadiceus* followed by interbreeding with *G. g. gallus* after their migration to Thailand several generations ago. Model-C fits the data best, with no f_4 outliers, unlike Model-A, Model-B and Model-D. All empirical *f*-statistics greater than 3 standard errors are defined as outliers.



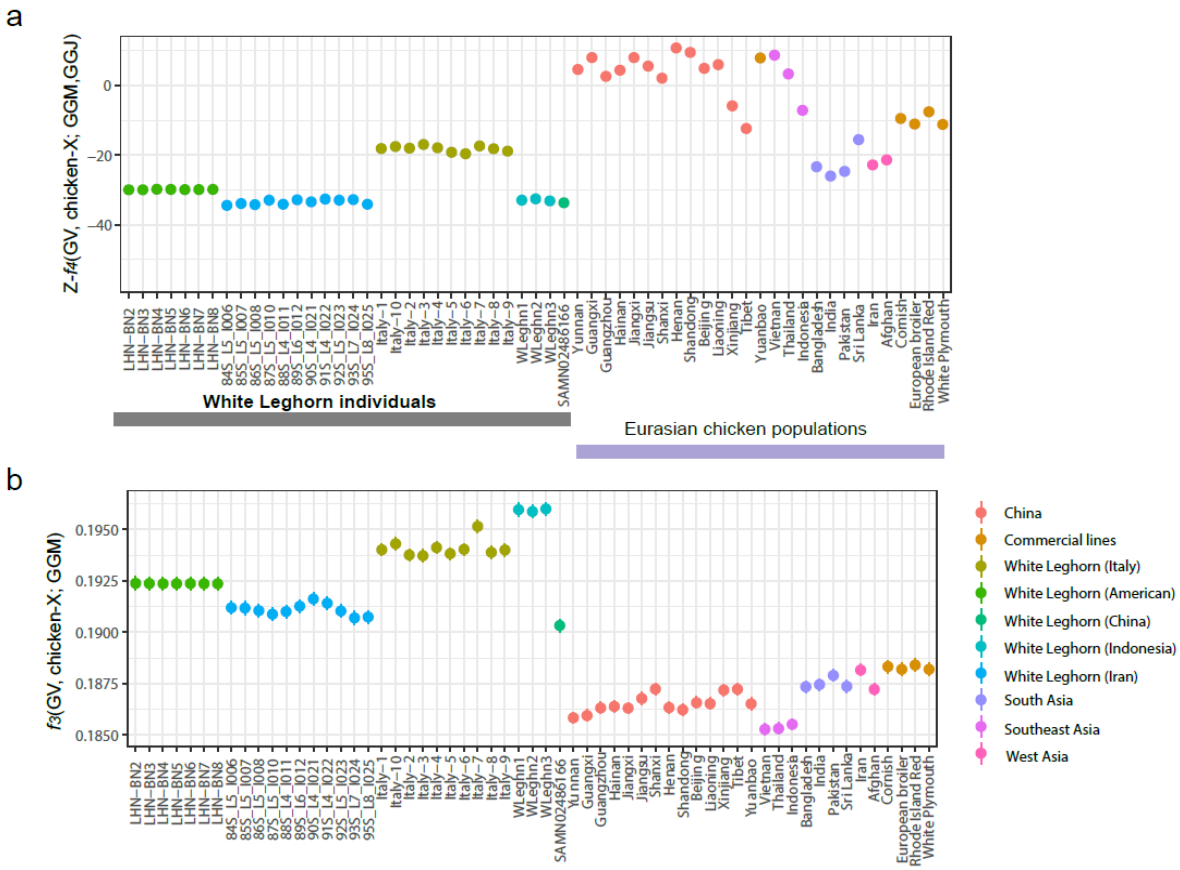
Supplementary information, Fig. S44.

f_4 statistics in form of f_4(GV, GGM; DC_X, WL) > 0 indicate that White Leghorn breed (WL) shares relatively higher genetic drift with *G. g. murghi* (GGM) compared with other domestic chicken populations. Despite that, f_4 statistics in form of f_4(GV, WL; DC_X, GGM) < 0 indicate that White Leghorn still shares comparatively higher genetic drift with other domestic chicken populations compared with *G. g. murghi*. GV, *G. varius*; DC_X, indigenous chickens grouped according to their sampling locations and also commercial lines as shown in the vertical axis.



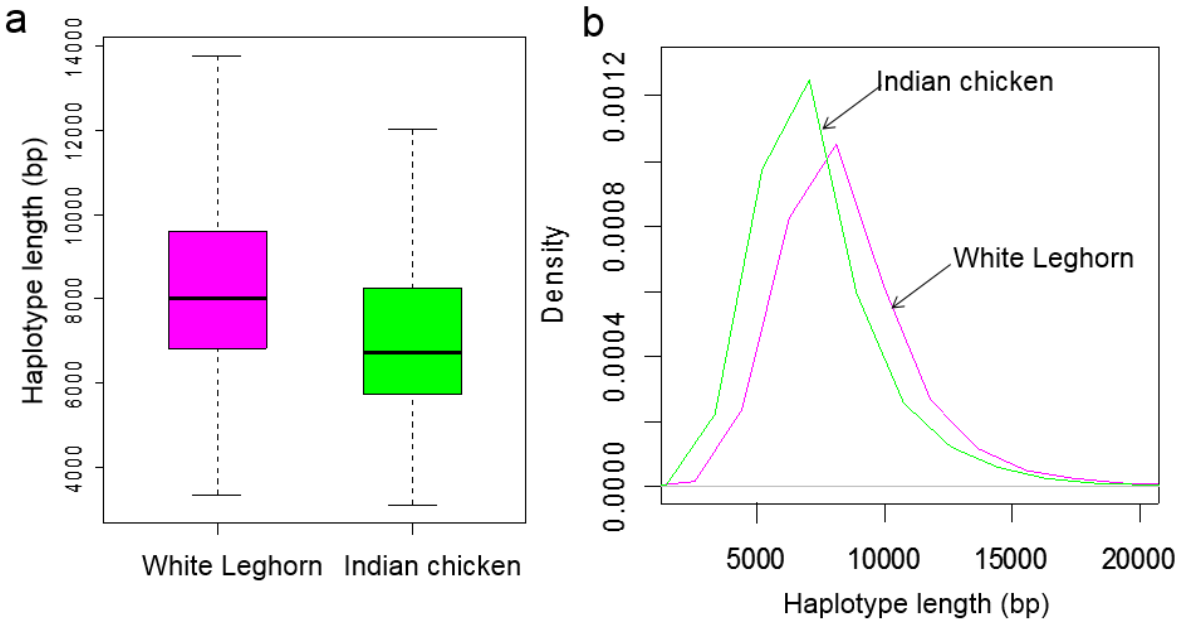
Supplementary information, Fig. S45.

White Leghorn breed carries a relatively higher level of *G. g. murghi* ancestry compared with other Eurasian chicken populations. **a** f_4 statistics show that White Leghorn is admixed with *G. g. murghi* (GGM) ($Z > 3$). **b** Outgroup-f_3 statistics demonstrate that White Leghorn carries a comparatively more ancestry from *G. g. murghi* compared with other Eurasian chicken populations. All *G. g. murghi* samples were included in this analysis.



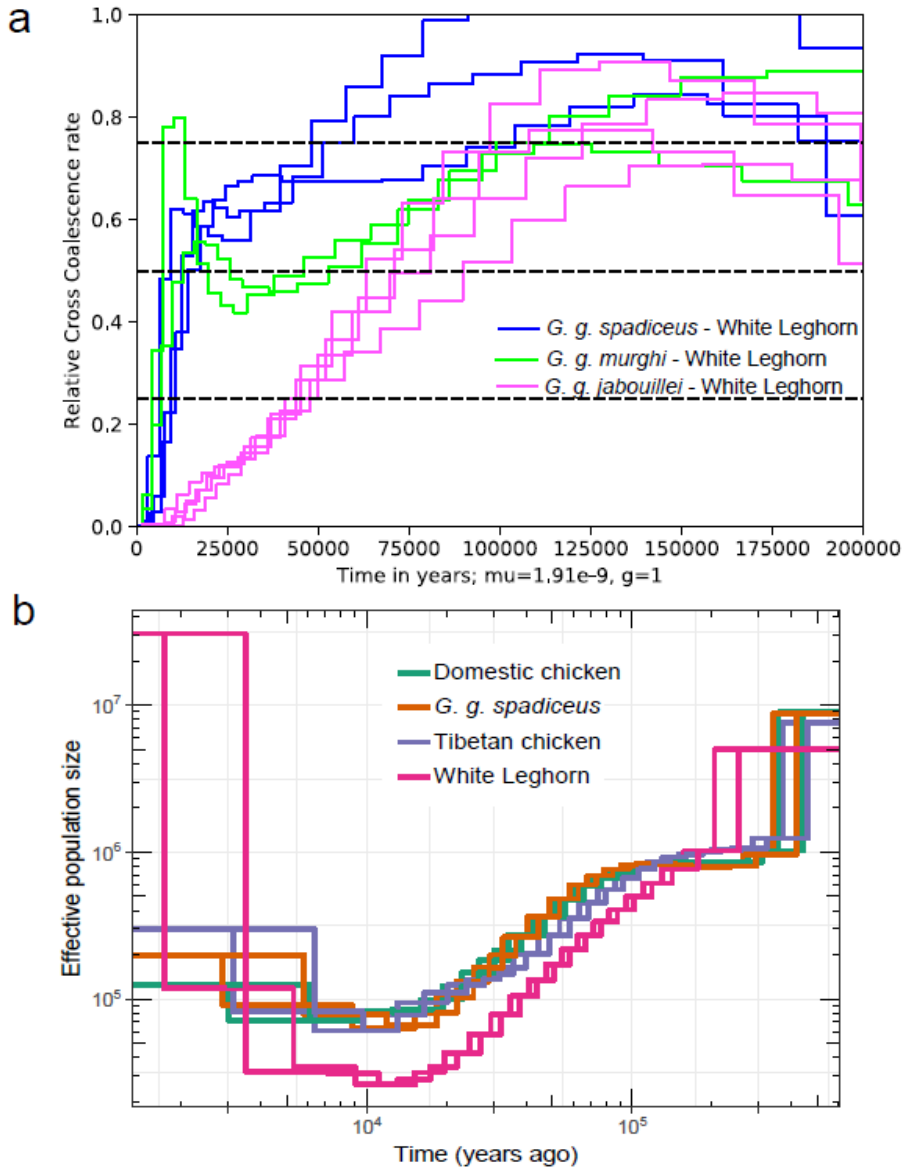
Supplementary information, Fig. S46.

Compared with other Eurasian chicken populations, the White Leghorn breed carries a comparatively high ancestry from *G. g. murghi* (GGM). **a** f_4 statistics show that White Leghorn was admixed with GGM ($Z < -3$). **b** Compared with other Eurasian chicken populations, outgroup-f_3 statistics show that White Leghorn carries relatively more ancestry from *G. g. murghi*. GGM used in these analyses excluded three admixed individuals with *G. g. spadiceus* which are identified from the results of early admixture and PCA analyses (Supplementary information, Figs. S21-S28). Interestingly, after excluding those three admixed birds, the outgroup-f_3 statistics indicate that the genetic affinity between GGM and White Leghorn becomes closer (larger f_3 values) compared with the result when all GGM were included in the analysis, further confirming the mosaic genomic ancestry of White Leghorn carrying a significantly high admixed component of GGM.



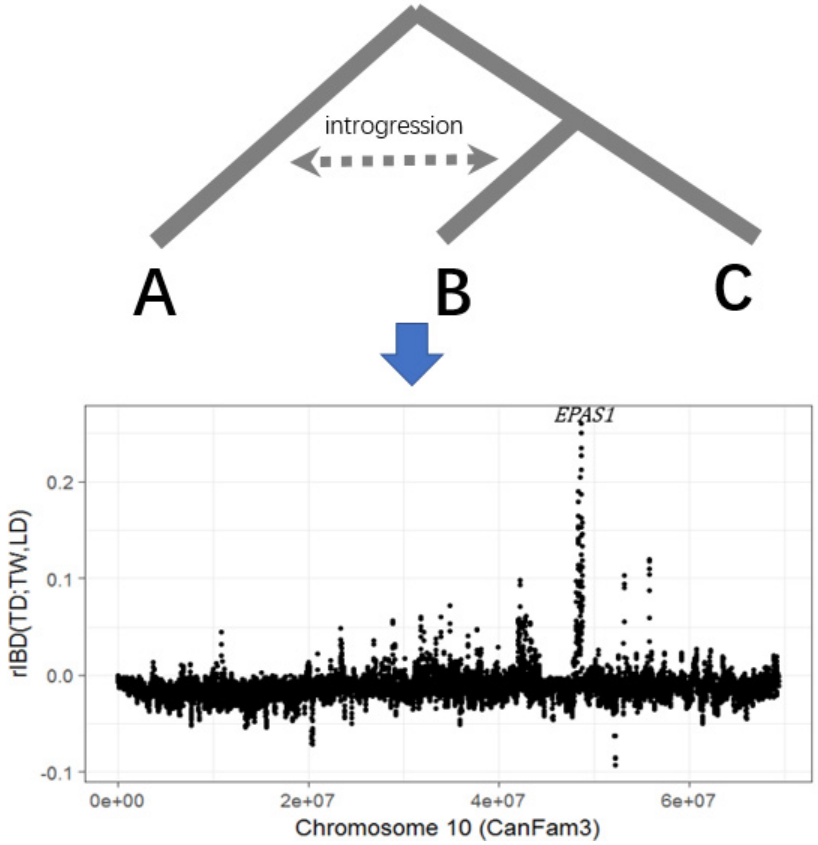
Supplementary information, Fig. S47.

Compared with Indian chickens, White Leghorn breed carries longer admixed genomic tracts from *G. g. murghi*. This observation suggests that White Leghorn and *G. g. murghi* have a relatively recent interbreeding event. **a** The haplotype length distribution of the admixed genomic tracts from *G. g. murghi* using box-plot. **b** The plots of haplotype length against density of the admixed genomic tracts from *G. g. murghi* in White Leghorn and Indian chickens. In this local ancestry inference by PCAdmix program (version: 1.0) [32], *G. g. murghi*, *G. g. spadiceus* and Yunnan chickens were used as source populations while Indian chickens and White Leghorn breed were treated as an 'admixed population' separately.



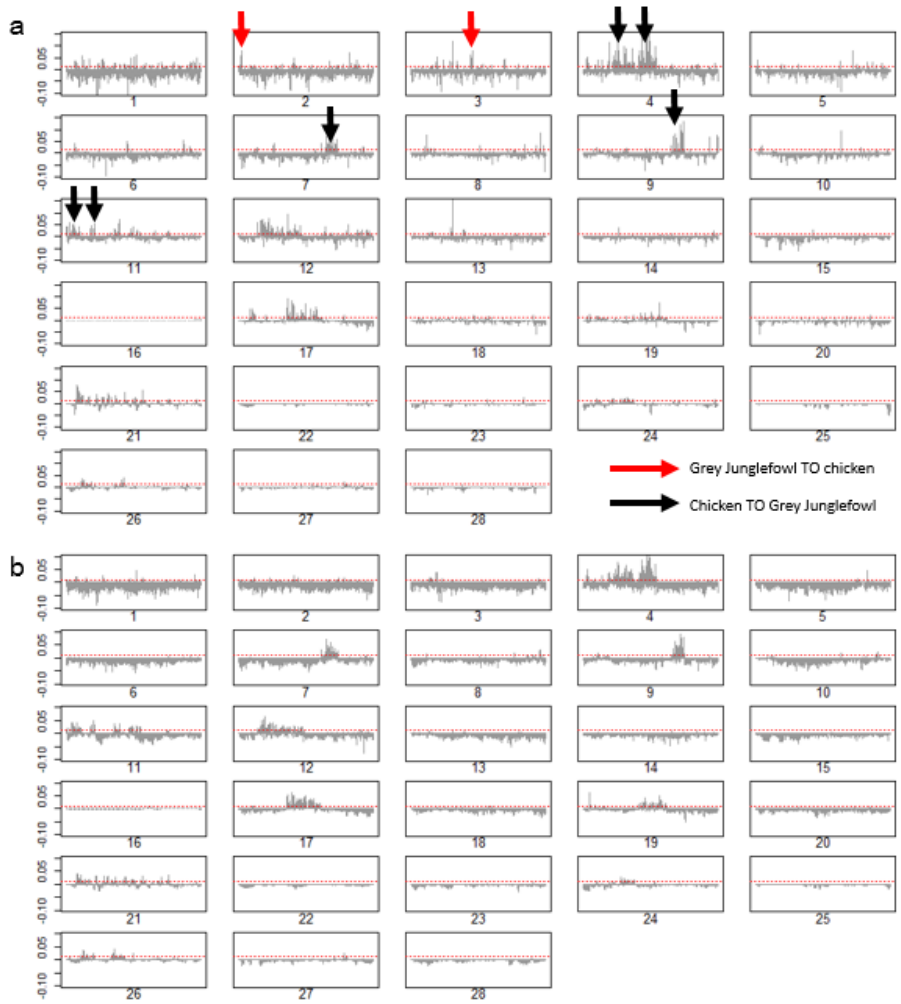
Supplementary information, Fig. S48.

MSMC analyses reveal *Ne* fluctuations and separation of White Leghorn breed from three Red Junglefowl subspecies. **a** White Leghorn shows a more recent coalescence with *G. g. spadiceus* than with other two Red Junglefowl subspecies, suggesting that White Leghorn was not derived from the other two Red Junglefowl subspecies, but from *G. g. spadiceus*. **b** MSMC analyses suggest a significant increase in recent *Ne* in the ancestral population of White Leghorn since around 3 kya, reflecting that the ancestral population of White Leghorn has a recent interbreeding history.



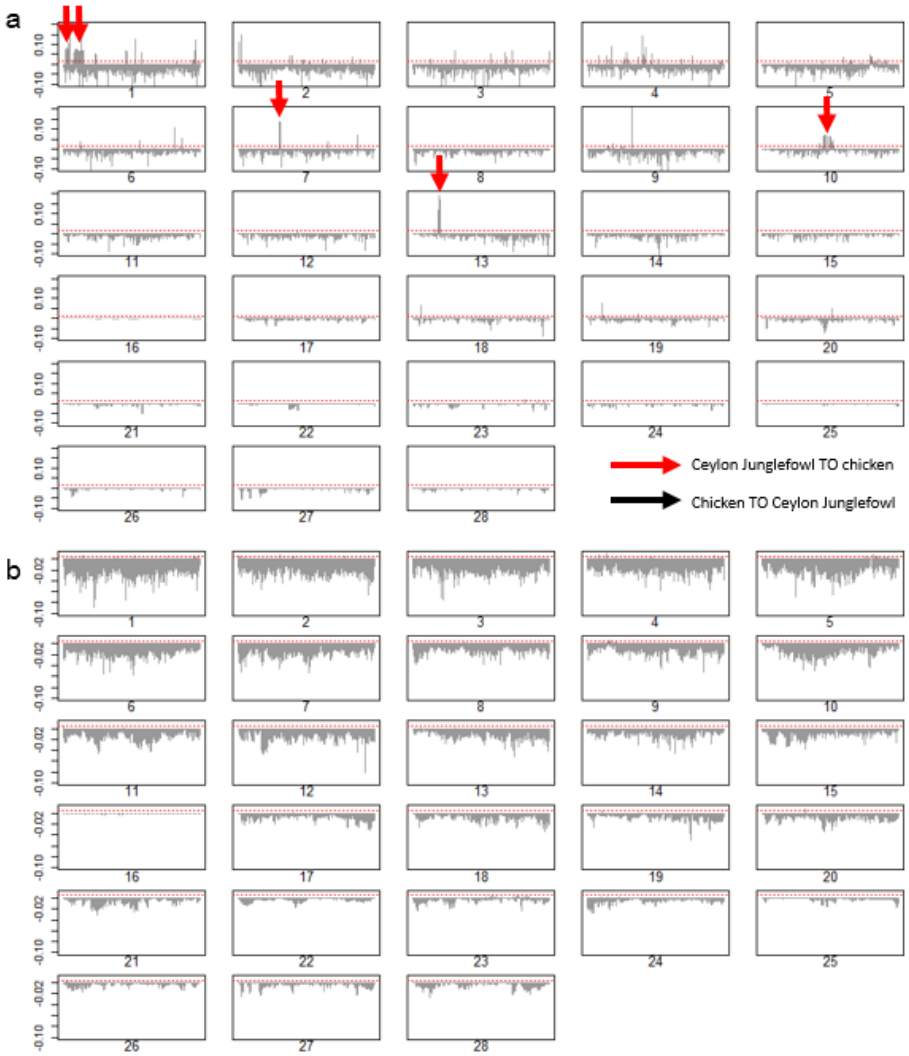
Supplementary information, Fig. S49.

An introduction for rIBD used for detecting the introgressed genomic regions. The top topology shows the relationship used for identifying introgressed IBD tracts in population A and B relative to population C, which is calculated in the form of rIBD(A;B,C). The point figure below indicates that the rIBD (TD;TW,LD) successfully replicates *EPAS1* gene and flanking region possessing the strongest signature of introgression in TW and TD as previously reported [49,50], suggesting the high power and efficiency of this method used in our study. TW, Tibetan gray wolf; TD, Tibetan dog; and LD, lowland dog.



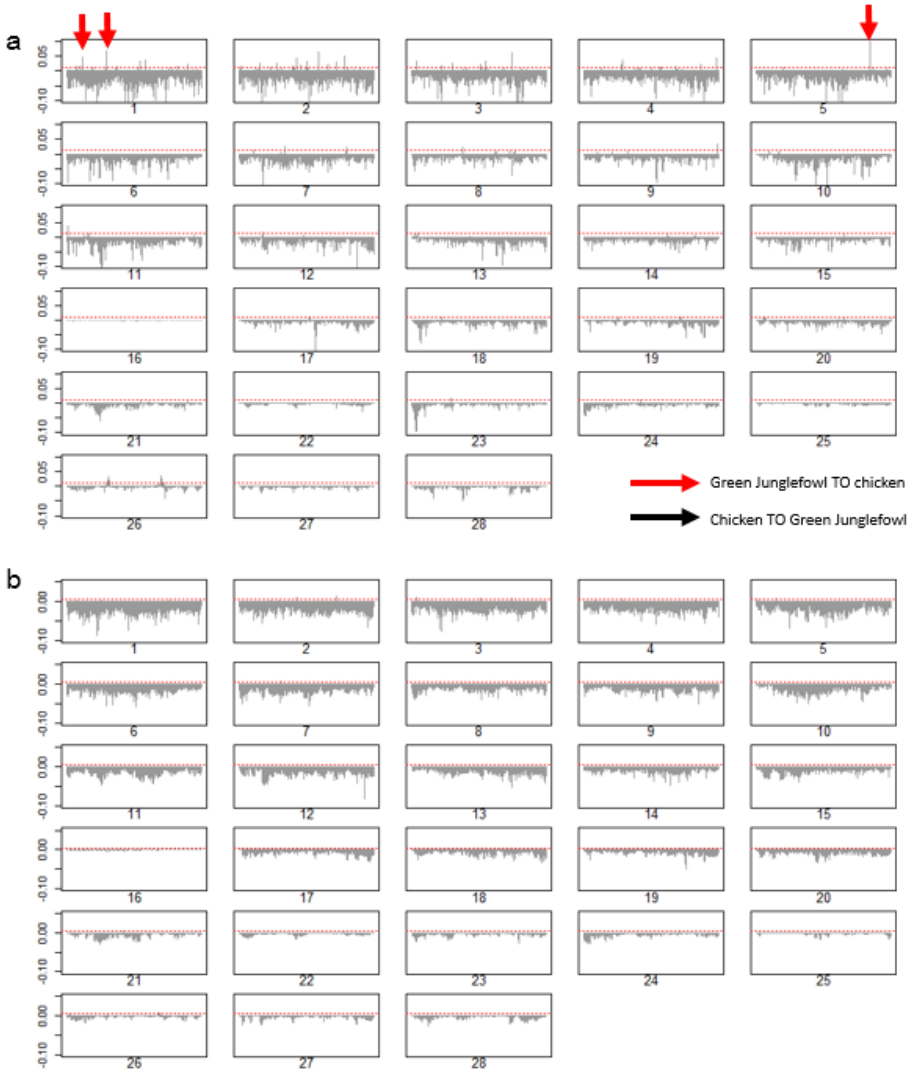
Supplementary information, Fig. S50.

Distribution of haplotypes across the whole genome shared by (**a**) Indian chickens and (**b**) all domestic chickens with Grey Junglefowl. The x-axis shows the full length of all autosomes while y-axis represents the relative frequency of Grey Junglefowl haplotypes to domestic chicken haplotypes compared with *G. g. spadiceus*, ranging from 1 (all haplotypes are IBD with Grey Junglefowl) to -1 (all haplotypes are IBD with *G. g. spadiceus*). Colored arrows depict the introgressed genomic regions and their directions.



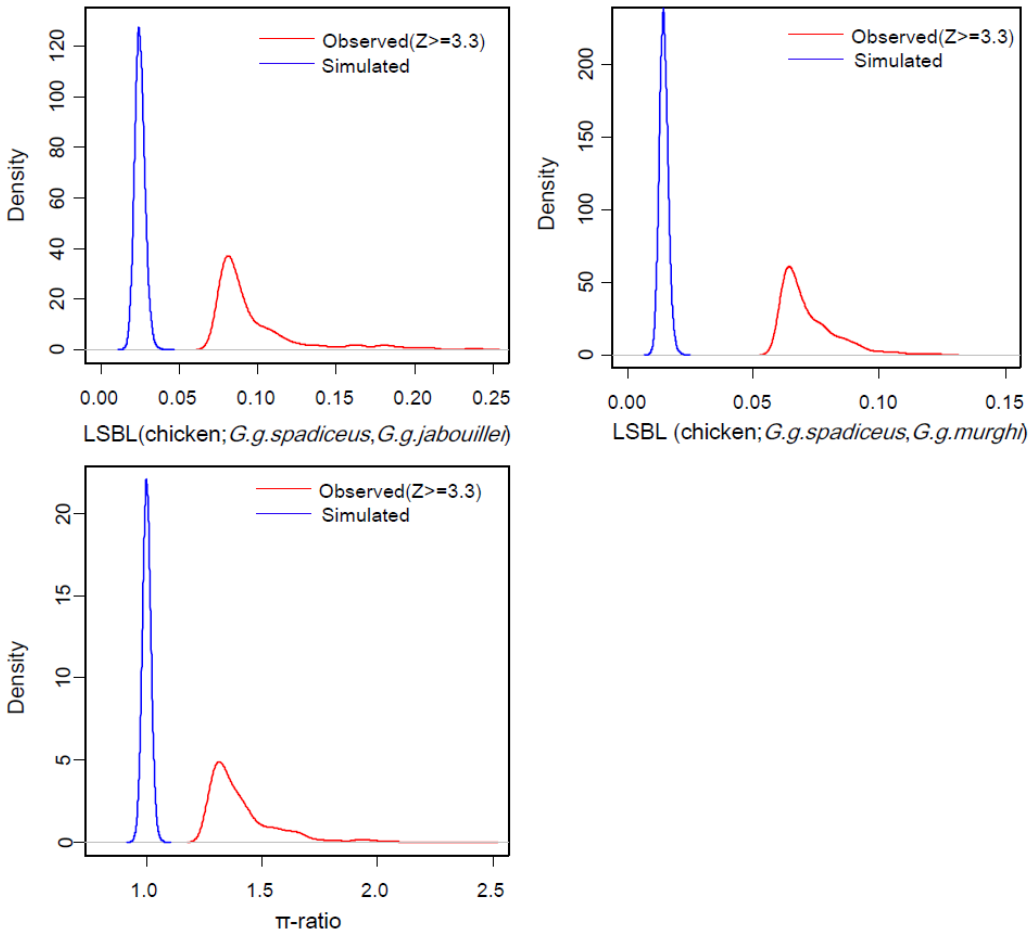
Supplementary information, Fig. S51.

Distribution of haplotypes shared by (**a**) Sri Lankan chickens and (**b**) all domestic chickens with the Ceylon Junglefowl. The x-axis shows the full length of all autosomes while the y-axis represents the relative frequency of Ceylon Junglefowl haplotypes to domestic chicken haplotypes compared with *G. g. spadiceus*, ranging from 1 (all haplotypes are IBD with Ceylon Junglefowl) to -1 (all haplotypes are IBD with *G. g. spadiceus*).



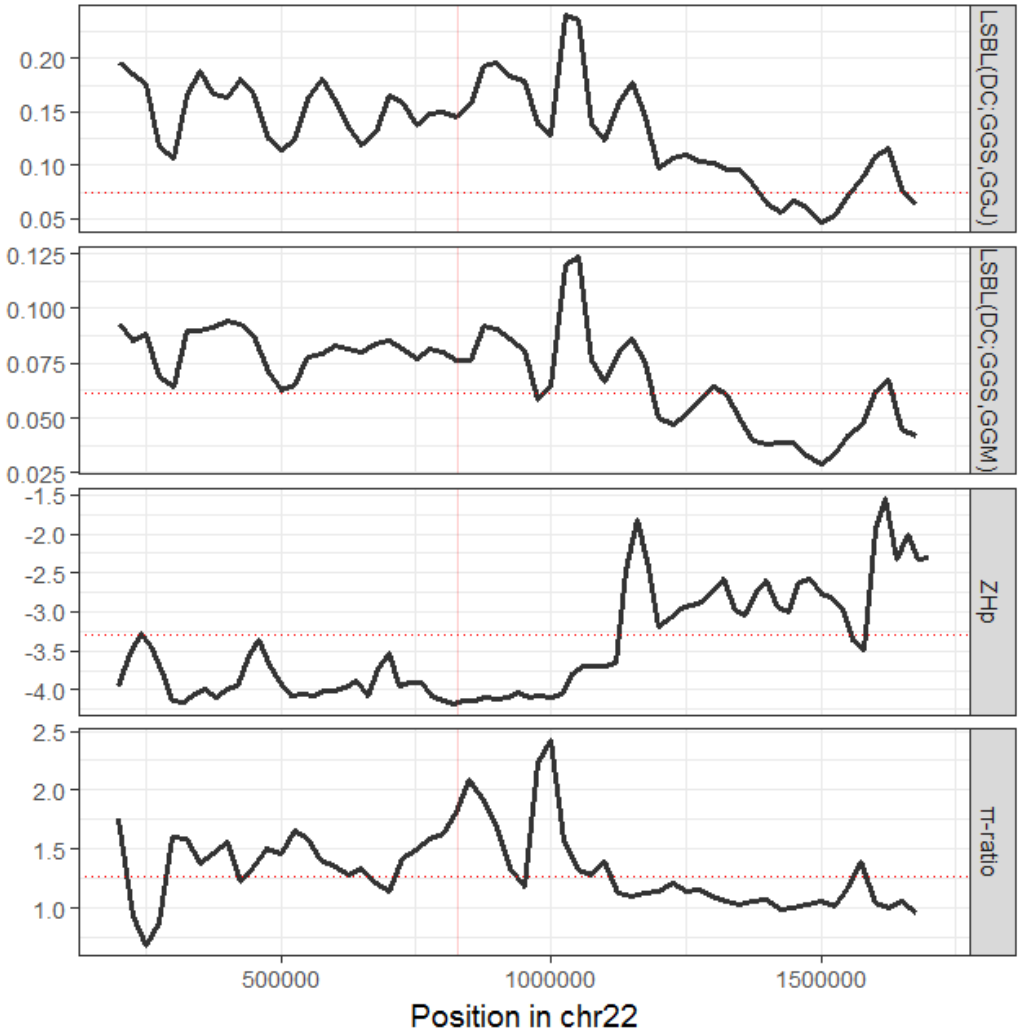
Supplementary information, Fig. S52.

Distribution of haplotypes shared by (**a**) Indonesian chickens and (**b**) all domestic chickens with the Green Junglefowl. The x-axis shows the full length of all autosomes and the y-axis represents the relative frequency of Green Junglefowl haplotypes to domestic chicken haplotypes compared with *G. g. spadiceus*, ranging from 1 (all haplotypes are IBD with Green Junglefowl) to -1 (all haplotypes are IBD with *G. g. spadiceus*). The two longest regions of consecutive introgressions are marked with arrows.



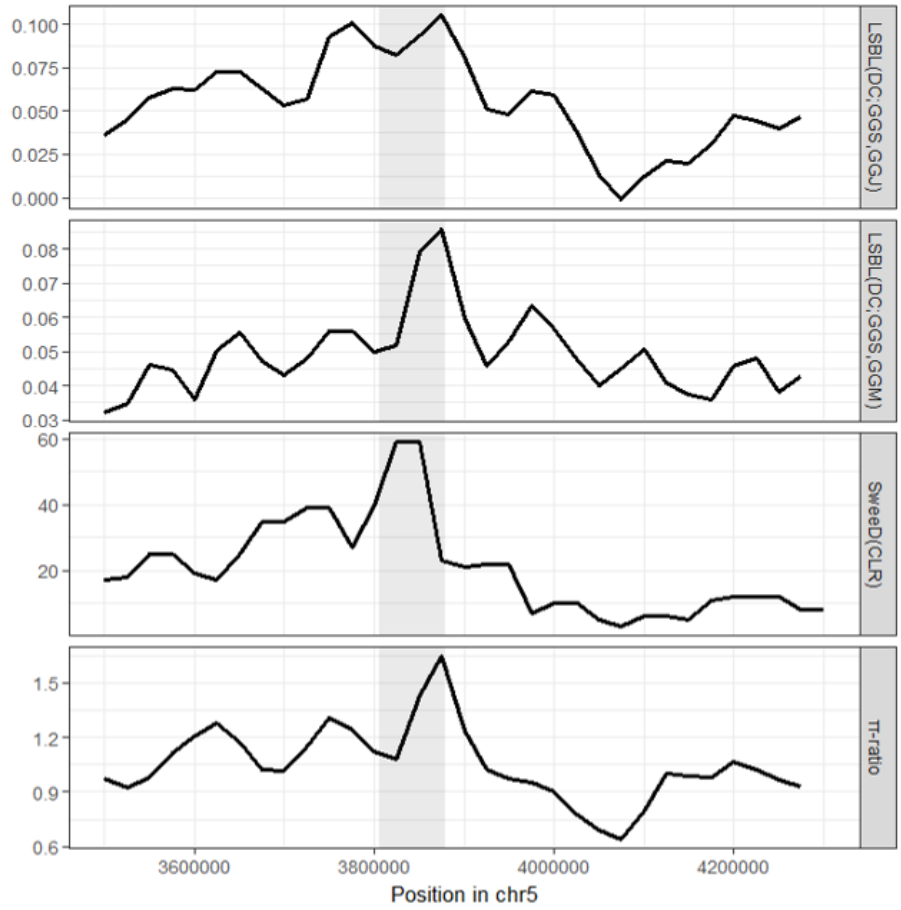
Supplementary information, Fig. S53.

The comparison of observed LSBL and π-ratio values (Z>3.3) with that calculated from simulated neutral sequences.



Supplementary information, Fig. S54.

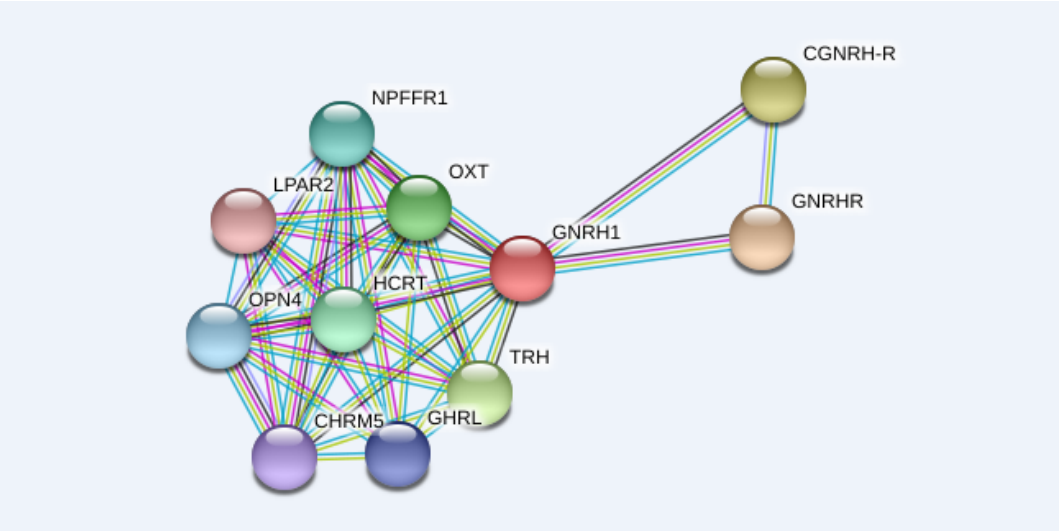
Selection signatures on gene *GNRH-1* was supported by each of four approaches.



Supplementary information, Fig. S55.

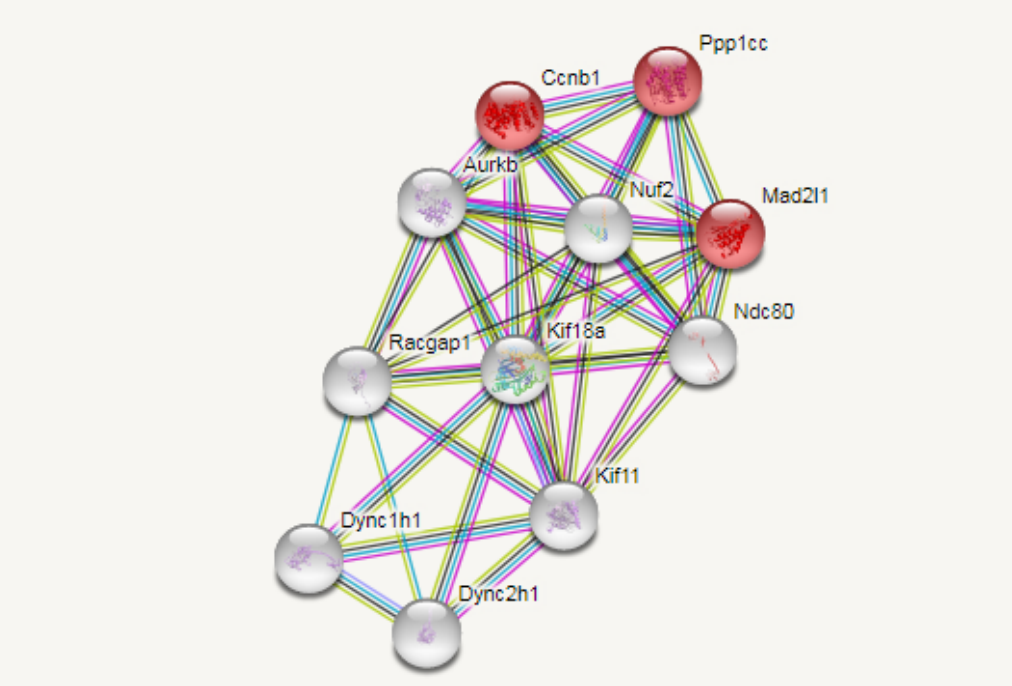
Selection signatures on gene *KIF18A* was supported by each of four approaches.

5



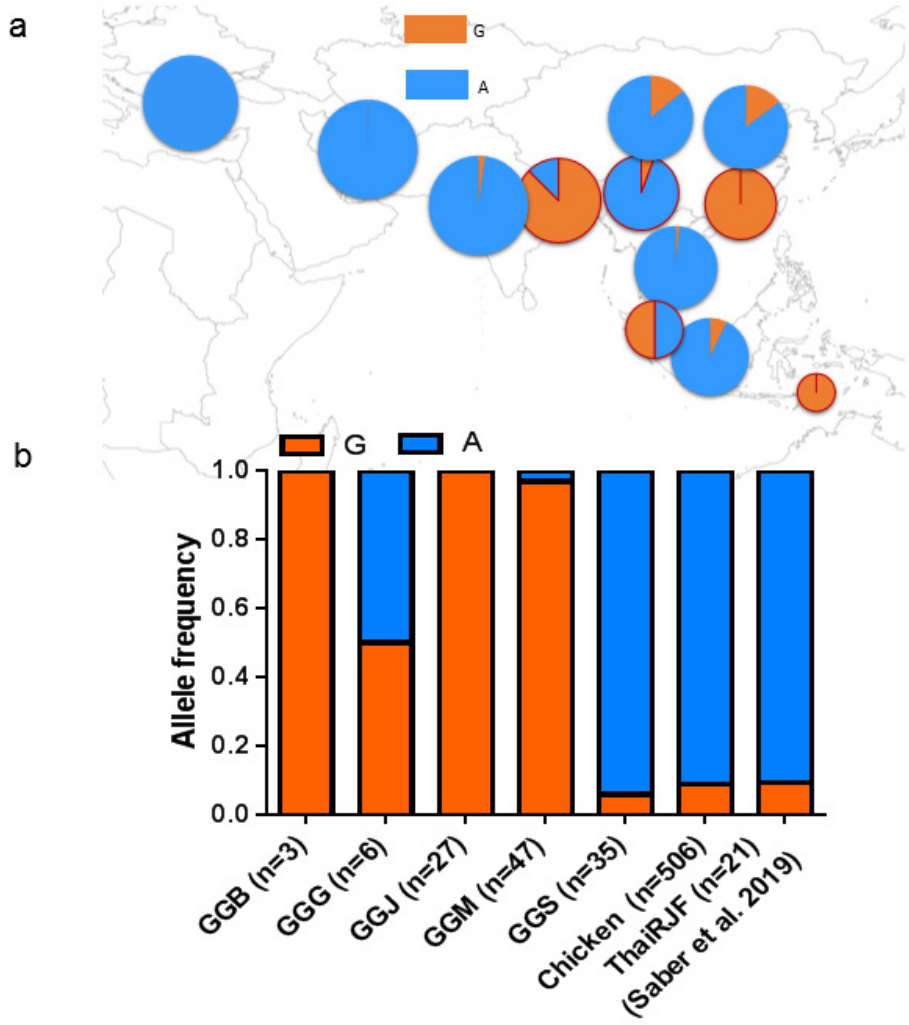
Supplementary information, Fig. S56.

Protein-protein interacted (PPI) network analysis of *GNRH1*. This analysis shows that GnRH-I is significantly enriched in a protein-protein interaction (PPI) network regulating reproduction (*P*=0.0008, adjusted by FDR). In the network, the gene is assigned to a key node directly connecting with *CGNRH-R* and *GNRHR*, both are known for their role in regulating early onset of puberty [79].



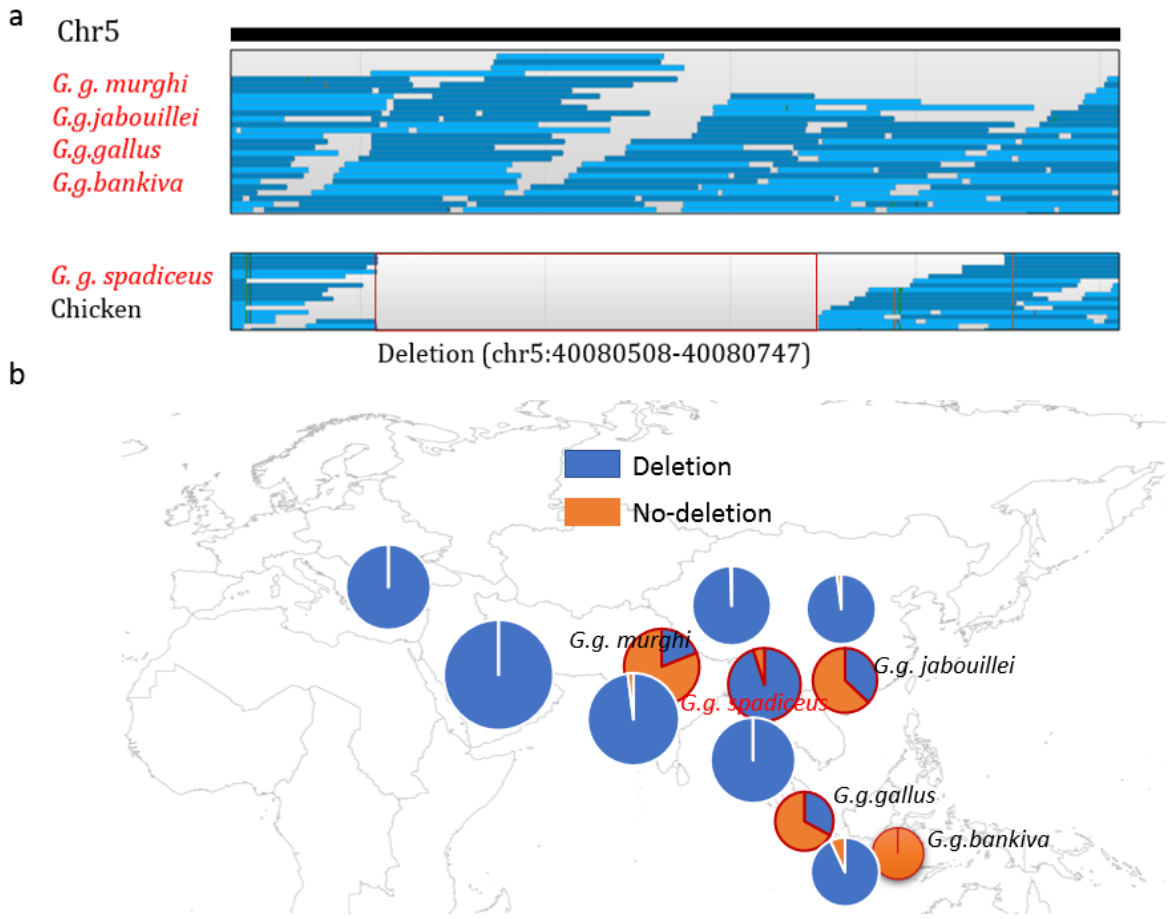
Supplementary information, Fig. S57.

Protein-protein interacted (PPI) network analysis of *KIF18A*. This analysis suggests that *KIF18A* is significantly connected with genes involved in oocyte meiosis processes (*P*=0.00428, adjusted by FDR).



Supplementary information, Fig. S58.

a Distribution of TSHR-Gly558Arg in domestic chickens and RJFs (pie with red frame). **b** Histogram denotes the overall frequencies in all domestic chickens and five RJF subspecies. GGS (*G. g. spadiceus*) is treated as the direct wild ancestor of domestic chickens. We indicated that the *TSHR* gene evolved under positive selection in both chickens and *G. g. spadiceus* and the TSHR-Gly558Arg mutation is nearly fixed in chickens, *G. g. spadiceus* and unclassified Thai RJFs reported by [69]. A half of *G. g. gallus* carried this mutation, which is possibly due to *G. g. gallus* has admixed with *G. g. spadiceus* and chicken (Supplementary information, Figs. S16, 36).



Supplementary information, Fig. S59.

a An overview of TSHR-deletion (chr5:40,080,508-40,080,747) in domestic chicken and Red junglefowl using Savant V2.05 (Fiume et al. Nucleic Acids Res. 2010, 26(16):1938-44). **b** Distribution of this deletion in domestic chickens and RJFs (pie with red frame).

Supplementary information, Table S1.

Information for samples and genome sequencing data used in this study (an Excel file).

Supplementary information, Table S2.

f_4 statistics in form of f_4(outgroup, A; B, C) show that *G. g. gallus* was admixed with both *G. g. spadiceus* and *G. g. jabouillei*. The Green Junglefowl was used as an outgroup. *G. g. gallus* and *G. g. murghi* have no overlap in their distribution ranges (Supplementary information, Fig. S1).

Outgroup	**A**	**B**	**C**	f_4	**SE**	**Z**
Green Junglefowl	*G. g. gallus*	*G. g. murghi*	*G. g. jabouillei*	0.000536	7.228E-05	7.4099
Green Junglefowl	*G. g. gallus*	*G. g. spadiceus*	*G. g. murghi*	-0.00401	8.768E-05	-45.71

Supplementary information, Table S3.

f_4 statistics in form of f_4(outgroup, A; B, C) show that two Thailand chickens (Thailand-2) and nine Thailand RJF (ThaiRJF) as mentioned in Fig.2a have more genetic diversity from domestic chickens than *G. g. spadiceus*, suggesting that both Thailand-2 and ThaiRJF were admixed with domestic chickens. Green Junglefowl was used as an outgroup.

Outgroup	A	B	C	f_4	SE	Z-score
Green Junglefowl	Thailand-2	*G. g. spadiceus*	Yunnan	0.00137831	6.75E-05	20.4176
Green Junglefowl	Thailand	*G. g. spadiceus*	Yunnan	0.0015154	7.14E-05	21.2117
Green Junglefowl	ThaiRJF	*G. g. spadiceus*	Yunnan	0.00172797	6.28E-05	27.5279

Supplementary information, Table S4.

Information for White Leghorn chickens sampled from Italy, Indonesia, and America included in this study for the investigation whether White Leghorn carries excess of ancestry from *G. g. murghi.*

Samples ID	Sampling location	Coverage	Note
Italy-1	Italy	7.97608	Newly generated
Italy-10	Italy	9.69965	Newly generated
Italy-2	Italy	7.89719	Newly generated
Italy-3	Italy	8.09881	Newly generated
Italy-4	Italy	8.85591	Newly generated
Italy-5	Italy	10.5259	Newly generated
Italy-6	Italy	10.4446	Newly generated
Italy-7	Italy	2.93903	Newly generated
Italy-8	Italy	12.5116	Newly generated
Italy-9	Italy	9.26773	Newly generated
WLeghn1	Indonesia	38.3549	Reference [80]
WLeghn2	Indonesia	38.6462	Reference [80]
WLeghn3	Indonesia	43.2905	Reference [80]
LHN-BN2	America	48.74	Kindly provided by Prof. Susan J. Lamont's Lab in Iowa State University
LHN-BN3	America	48.79	Kindly provided by Prof. Susan J. Lamont's Lab in Iowa State University
LHN-BN4	America	42.39	Kindly provided by Prof. Susan J. Lamont's Lab in Iowa State University
LHN-BN5	America	43.82	Kindly provided by Prof. Susan J. Lamont's Lab in Iowa State University
LHN-BN6	America	62.33	Kindly provided by Prof. Susan J. Lamont's Lab in Iowa State University
LHN-BN7	America	46.07	Kindly provided by Prof. Susan J. Lamont's Lab in Iowa State University
LHN-BN8	America	46.36	Kindly provided by Prof. Susan J. Lamont's Lab in Iowa State University

Supplementary information, Table S5.

D-statistic in form of *D*(Green Junglefowl, *G. g. murghi*; POP_Y, Cornish) indicate Cornish, contains a proportion of *G. g. murghi* ancestry. $|Z| > 3$ is considered to be statistically significant.

Pop(W)	Pop(X)	Pop(Y)	Pop(Z)	D-value	Z-value	No. of SNP	No. of blocks
Green Junglefowl	*G. g. murghi*	Afghanistan	Cornish	0.0046	3.338116268	115201	1918601
Green Junglefowl	*G. g. murghi*	Beijing, China	Cornish	0.0113	7.829110865	108386	1918570
Green Junglefowl	*G. g. murghi*	Bangladesh	Cornish	0.0031	2.31211639	115674	1918601
Green Junglefowl	*G. g. murghi*	India	Cornish	0.0035	2.710115275	114481	1918601
Green Junglefowl	***G. g. murghi***	**White Leghorn**	**Cornish**	**-0.018**	**-10.67310689**	**110795**	**1918601**
Green Junglefowl	*G. g. murghi*	European broiler	Cornish	0.0005	0.334102238	102143	1916222
Green Junglefowl	*G. g. murghi*	Guangxi, China	Cornish	0.0159	11.33711758	113885	1918579
Green Junglefowl	*G. g. murghi*	Guangdong, China	Cornish	0.0137	10.23511255	109518	1917299
Green Junglefowl	*G. g. murghi*	Hainan, China	Cornish	0.0147	9.51611388	110571	1916098
Green Junglefowl	*G. g. murghi*	Henan, China	Cornish	0.0144	8.976112231	109051	1918494
Green Junglefowl	*G. g. murghi*	Indonesia	Cornish	0.0188	14.62711787	113519	1918601
Green Junglefowl	*G. g. murghi*	Iran	Cornish	-0.0032	-2.481109399	110098	1918601
Green Junglefowl	*G. g. murghi*	Jiangsu, China	Cornish	0.0104	6.471110448	108164	1918501
Green Junglefowl	*G. g. murghi*	Jiangxi, China	Cornish	0.013	9.7701129	109999	1918601
Green Junglefowl	*G. g. murghi*	Liaoning, China	Cornish	0.0117	8.066110803	108233	1918588
Green Junglefowl	*G. g. murghi*	Pakistan	Cornish	-0.0009	-0.683112963	113168	1918601
Green Junglefowl	*G. g. murghi*	Rhode Island Red	Cornish	0.0015	0.793100304	100001	1901506
Green Junglefowl	*G. g. murghi*	Shandong, China	Cornish	0.0137	10.0741128	109745	1918601
Green Junglefowl	*G. g. murghi*	Shanxi, China	Cornish	0.0061	3.973109317	108001	1918554
Green Junglefowl	*G. g. murghi*	Sri Lanka	Cornish	0.0052	3.153111818	110668	1918590
Green Junglefowl	*G. g. murghi*	Taiwan, China	Cornish	0.002	1.05810796	107533	1917895
Green Junglefowl	*G. g. murghi*	Thailand	Cornish	0.0212	14.82911926	114304	1918539
Green Junglefowl	*G. g. murghi*	ThaiRJF	Cornish	0.0208	15.21812031	115407	1918601
Green Junglefowl	*G. g. murghi*	Thailand-2	Cornish	0.0277	18.32912276	116152	1916924
Green Junglefowl	*G. g. murghi*	Tibet, China	Cornish	0.005	3.76311703	115867	1918601
Green Junglefowl	*G. g. murghi*	Vietnam	Cornish	0.0214	14.25811899	114006	1918593
Green Junglefowl	*G. g. murghi*	White Plymouth	Cornish	-0.0004	-0.207100905	100977	1917647
Green Junglefowl	*G. g. murghi*	Xinjiang, China	Cornish	0.0056	4.412111156	109909	1918601
Green Junglefowl	*G. g. murghi*	Yuanbao, China	Cornish	0.0115	7.089112638	110082	1918601
Green Junglefowl	*G. g. murghi*	Yunnan, China	Cornish	0.0161	11.90011774	114000	1918601
Green Junglefowl	*G. g. murghi*	*G. g. spadiceus*	Cornish	0.0172	12.85611972	115672	1918601

5

Supplementary information, Table S6.

Compared with other chicken populations, *D*-statistics in form of *D*(Green Junglefowl, *G. g. murghi*; POP_Y, White Leghorn) indicate that White Leghorn contains more *G. g. murghi* ancestry. $|Z| > 3$ is considered to be statistically significant.

Pop(W)	Pop(X)	Pop(Y)	Pop(Z)	D-value	Z-value	No. of SNP	No. of blocks
Green junglefowl	*G. g. murghi*	Afghanistan	White Leghorn	0.0219	16.83511616	111182	1925699
Green junglefowl	*G. g. murghi*	Beijing, China	White Leghorn	0.0278	20.3391186	112191	1925667
Green junglefowl	*G. g. murghi*	Bangladesh	White Leghorn	0.0205	16.7291152	110577	1925699
Green junglefowl	*G. g. murghi*	Cornish	White Leghorn	0.018	10.6731108	106886	1918601
Green junglefowl	*G. g. murghi*	India	White Leghorn	0.0209	17.14411512	110407	1925699
Green junglefowl	*G. g. murghi*	European broiler	White Leghorn	0.0185	14.57211043	106421	1923117
Green junglefowl	*G. g. murghi*	Guangxi, China	White Leghorn	0.032	24.00012304	115415	1925675
Green junglefowl	*G. g. murghi*	Guangdong, China	White Leghorn	0.03	21.75011969	112713	1924383
Green junglefowl	*G. g. murghi*	Hainan, China	White Leghorn	0.0311	21.55612025	112999	1922935
Green junglefowl	*G. g. murghi*	Henan, China	White Leghorn	0.0301	19.97312132	114217	1925586
Green junglefowl	*G. g. murghi*	Indonesia	White Leghorn	0.0359	30.66311941	111131	1925699
Green junglefowl	*G. g. murghi*	Iran	White Leghorn	0.0149	14.04610975	106538	1925699
Green junglefowl	*G. g. murghi*	Jiangsu, China	White Leghorn	0.0269	18.30911861	112402	1925585
Green junglefowl	*G. g. murghi*	Jiangxi, China	White Leghorn	0.0291	22.47812077	113938	1925699
Green junglefowl	*G. g. murghi*	Liaoning, China	White Leghorn	0.0281	20.70711906	112558	1925685
Green junglefowl	*G. g. murghi*	Pakistan	White Leghorn	0.017	15.85411067	106966	1925699
Green junglefowl	*G. g. murghi*	Rhode Island Red	White Leghorn	0.0188	11.68111225	108098	1907684
Green junglefowl	*G. g. murghi*	Shandong, China	White Leghorn	0.0297	22.36912109	114097	1925699
Green junglefowl	*G. g. murghi*	Shanxi, China	White Leghorn	0.0231	16.92611604	110807	1925646
Green junglefowl	*G. g. murghi*	Sri Lanka	White Leghorn	0.0231	16.31711205	106979	1925685
Green junglefowl	*G. g. murghi*	Taiwan, China	White Leghorn	0.0199	11.51911121	106883	1924915
Green junglefowl	*G. g. murghi*	Thailand	White Leghorn	0.0371	26.68012424	115361	1925636
Green junglefowl	*G. g. murghi*	ThaiRJF	White Leghorn	0.0367	27.91412477	115941	1925699
Green junglefowl	*G. g. murghi*	Thailand-2	White Leghorn	0.0432	30.24012736	116820	1923957
Green junglefowl	*G. g. murghi*	Tibet, China	White Leghorn	0.0222	18.52011676	111675	1925699
Green junglefowl	*G. g. murghi*	Vietnam	White Leghorn	0.0371	26.84712478	115861	1925689
Green junglefowl	*G. g. murghi*	White Plymouth	White Leghorn	0.0177	10.99911062	106773	1924522
Green junglefowl	*G. g. murghi*	Xinjiang, China	White Leghorn	0.023	19.31711504	109868	1925699
Green junglefowl	*G. g. murghi*	Yuanbao, China	White Leghorn	0.0275	18.56912093	114449	1925699
Green junglefowl	*G. g. murghi*	Yunnan, China	White Leghorn	0.0322	24.63712293	115263	1925699
Green junglefowl	*G. g. murghi*	*G. g. spadiceus*	White Leghorn	0.0333	27.15712358	115605	1925699

Supplementary information, Table S7.

D-statistics in form of *D*(Green Junglefowl, *G. g. murghi*; POP_Y, European broiler) indicate that European broiler contains some *G. g. murghi* ancestry. The proportion of *G. g. murghi* ancestry is higher in White Leghorn than in European broiler. $|Z| > 3$ is considered to be statistically significant.

Pop(W)	Pop(X)	Pop(Y)	Pop(Z)	D-value	Z-value	No. of SNP	No. of blocks
Green Junglefowl	*G. g. murghi*	Afghanistan	European broiler	0.0042	4.273116364	115399	1923117
Green Junglefowl	*G. g. murghi*	Beijing, China	European broiler	0.0109	10.71311126	108861	1923085
Green Junglefowl	*G. g. murghi*	Bangladesh	European broiler	0.0027	3.097116329	115711	1923117
Green Junglefowl	*G. g. murghi*	Cornish	European broiler	-0.0005	-0.334102143	102238	1916222
Green Junglefowl	*G. g. murghi*	India	European broiler	0.003	3.51211538	114680	1923117
Green Junglefowl	***G. g. murghi***	**White Leghorn**	**European broiler**	**-0.0185**	**-14.57210642**	**110431**	**1923117**
Green Junglefowl	*G. g. murghi*	Guangxi, China	European broiler	0.0156	16.73011769	114075	1923093
Green Junglefowl	*G. g. murghi*	Guangdong, China	European broiler	0.0133	14.03311302	110062	1921801
Green Junglefowl	*G. g. murghi*	Hainan, China	European broiler	0.0143	12.25611437	111136	1920434
Green Junglefowl	*G. g. murghi*	Henan, China	European broiler	0.0139	11.72311251	109412	1923013
Green Junglefowl	*G. g. murghi*	Indonesia	European broiler	0.0184	22.04911788	113611	1923117
Green Junglefowl	*G. g. murghi*	Iran	European broiler	-0.0036	-4.907109144	109941	1923117
Green Junglefowl	*G. g. murghi*	Jiangsu, China	European broiler	0.0099	9.055111188	109004	1923031
Green Junglefowl	*G. g. murghi*	Jiangxi, China	European broiler	0.0126	13.68711325	110427	1923117
Green Junglefowl	*G. g. murghi*	Liaoning, China	European broiler	0.0113	11.03411116	108666	1923103
Green Junglefowl	*G. g. murghi*	Pakistan	European broiler	-0.0013	-1.66611281	113109	1923117
Green Junglefowl	*G. g. murghi*	Rhode Island Red	European broiler	0.0008	0.622107144	106962	1905459
Green Junglefowl	*G. g. murghi*	Shandong, China	European broiler	0.0134	14.84511318	110196	1923117
Green Junglefowl	*G. g. murghi*	Shanxi, China	European broiler	0.0056	5.395109695	108473	1923066
Green Junglefowl	*G. g. murghi*	Sri Lanka	European broiler	0.0047	3.67511215	111094	1923105
Green Junglefowl	*G. g. murghi*	Taiwan, China	European broiler	0.0015	0.948108114	107793	1922335
Green Junglefowl	*G. g. murghi*	Thailand	European broiler	0.0208	19.29211949	114610	1923057
Green Junglefowl	*G. g. murghi*	ThaiRJF	European broiler	0.0205	21.66812044	115615	1923117
Green Junglefowl	*G. g. murghi*	Thailand-2	European broiler	0.0273	24.39412299	116457	1921395
Green Junglefowl	*G. g. murghi*	Tibet, China	European broiler	0.0046	5.276117053	115981	1923117
Green Junglefowl	*G. g. murghi*	Vietnam	European broiler	0.0211	20.00811927	114351	1923111
Green Junglefowl	*G. g. murghi*	White Plymouth	European broiler	-0.0008	-0.602101579	101748	1921977
Green Junglefowl	*G. g. murghi*	Xinjiang, China	European broiler	0.0052	6.71111127	110115	1923117
Green Junglefowl	*G. g. murghi*	Yuanbao, China	European broiler	0.011	9.738113252	110786	1923117
Green Junglefowl	*G. g. murghi*	Yunnan, China	European broiler	0.0158	18.06211802	114354	1923117
Green Junglefowl	*G. g. murghi*	*G. g. spadiceus*	European broiler	0.0168	19.16811993	115971	1923117

Supplementary information, Table S8.

D-statistics in form of *D*(Green Junglefowl, *G. g. murghi*; POP_Y, Rhode Island Red) indicate that Rhode Island Red contains some *G. g. murghi* ancestry but lower than that in White Leghorn. |Z| > 3 is considered to be statistically significant.

Pop(W)	Pop(X)	Pop(Y)	Pop(Z)	D-value	Z-value	No. of SNP	No. of blocks
Green Junglefowl	*G. g. murghi*	Afghanistan	Rhode Island Red	0.0032	2.487116625	115871	1907684
Green Junglefowl	*G. g. murghi*	Beijing, China	Rhode Island Red	0.0101	8.054109978	107784	1907652
Green Junglefowl	*G. g. murghi*	Bangladesh	Rhode Island Red	0.0018	1.614115753	115331	1907684
Green Junglefowl	*G. g. murghi*	Cornish	Rhode Island Red	-0.0015	-0.793100001	100304	1901506
Green Junglefowl	*G. g. murghi*	India	Rhode Island Red	0.0022	1.793115146	114634	1907684
Green Junglefowl	***G. g. murghi***	**White Leghorn**	**Rhode Island Red**	**-0.0188**	**-11.6811081**	**112245**	**1907684**
Green Junglefowl	*G. g. murghi*	European broiler	Rhode Island Red	-0.0008	-0.622106962	107144	1905459
Green Junglefowl	*G. g. murghi*	Guangxi, China	Rhode Island Red	0.0147	12.15311704	113642	1907661
Green Junglefowl	*G. g. murghi*	Guangdong, China	Rhode Island Red	0.0124	9.727112506	109753	1906397
Green Junglefowl	*G. g. murghi*	Hainan, China	Rhode Island Red	0.0135	9.593113572	110553	1905357
Green Junglefowl	*G. g. murghi*	Henan, China	Rhode Island Red	0.013	9.280111353	108486	1907581
Green Junglefowl	*G. g. murghi*	Indonesia	Rhode Island Red	0.0176	14.75711732	113266	1907684
Green Junglefowl	*G. g. murghi*	Iran	Rhode Island Red	-0.0046	-3.797107899	108886	1907684
Green Junglefowl	*G. g. murghi*	Jiangsu, China	Rhode Island Red	0.0091	6.896110007	108027	1907586
Green Junglefowl	*G. g. murghi*	Jiangxi, China	Rhode Island Red	0.0118	9.733112204	109592	1907684
Green Junglefowl	*G. g. murghi*	Liaoning, China	Rhode Island Red	0.0104	7.96310974	107482	1907672
Green Junglefowl	*G. g. murghi*	Pakistan	Rhode Island Red	-0.0022	-1.860110954	111449	1907684
Green Junglefowl	*G. g. murghi*	Shandong, China	Rhode Island Red	0.0125	10.11711184	109070	1907684
Green Junglefowl	*G. g. murghi*	Shanxi, China	Rhode Island Red	0.0049	3.60610533	104300	1907642
Green Junglefowl	*G. g. murghi*	Sri Lanka	Rhode Island Red	0.004	2.48710958	108717	1907673
Green Junglefowl	*G. g. murghi*	Taiwan, China	Rhode Island Red	0.0005	0.287108145	108038	1906958
Green Junglefowl	*G. g. murghi*	Thailand	Rhode Island Red	0.0199	14.88111889	114245	1907627
Green Junglefowl	*G. g. murghi*	ThaiRJF	Rhode Island Red	0.0196	16.18911961	115018	1907684
Green Junglefowl	*G. g. murghi*	Thailand-2	Rhode Island Red	0.0264	19.54412242	116134	1906063
Green Junglefowl	*G. g. murghi*	Tibet, China	Rhode Island Red	0.0038	3.20011653	115650	1907684
Green Junglefowl	*G. g. murghi*	Vietnam	Rhode Island Red	0.0202	15.58111863	113931	1907676
Green Junglefowl	*G. g. murghi*	White Plymouth	Rhode Island Red	-0.0018	-1.077106246	106629	1906711
Green Junglefowl	*G. g. murghi*	Xinjiang, China	Rhode Island Red	0.0043	3.558110345	109393	1907684
Green Junglefowl	*G. g. murghi*	Yuanbao, China	Rhode Island Red	0.0102	7.303111951	109701	1907684
Green Junglefowl	*G. g. murghi*	Yunnan, China	Rhode Island Red	0.0149	12.49711719	113745	1907684
Green Junglefowl	*G. g. murghi*	*G. g. spadiceus*	Rhode Island Red	0.0159	13.65611916	115424	1907684

Supplementary information, Table S9.

D-statistics in form of *D*(Green Junglefowl, *G. g. murghi*; POP_Y, White Plymouth) indicate that White Plymouth contains *G. g. murghi* ancestry. The proportion of *G. g. murghi* ancestry in White Leghorn is higher than in White Plymouth. |Z| >3 is considered to be statistically significant.

Pop(W)	Pop(X)	Pop(Y)	Pop(Z)	D-value	Z-value	No. of SNP	No. of blocks
Green Junglefowl	*G. g. murghi*	Afghanistan	White Plymouth	0.0049	3.340116759	115630	1924522
Green Junglefowl	*G. g. murghi*	Beijing, China	White Plymouth	0.0117	8.332110613	108051	1924491
Green Junglefowl	*G. g. murghi*	Bangladesh	White Plymouth	0.0034	2.479116526	115741	1924522
Green Junglefowl	*G. g. murghi*	Cornish	White Plymouth	0.0004	0.207100977	100905	1917647
Green Junglefowl	*G. g. murghi*	India	White Plymouth	0.0038	2.739115566	114700	1924522
Green Junglefowl	***G. g. murghi***	**White Leghorn**	**White Plymouth**	**-0.0177**	**-10.99910677**	**110616**	**1924522**
Green Junglefowl	*G. g. murghi*	European broiler	White Plymouth	0.0008	0.602101748	101579	1921977
Green Junglefowl	*G. g. murghi*	Guangxi, China	White Plymouth	0.0164	12.09911732	113544	1924500
Green Junglefowl	*G. g. murghi*	Guangdong, China	White Plymouth	0.0142	9.612112013	108887	1923220
Green Junglefowl	*G. g. murghi*	Hainan, China	White Plymouth	0.0152	9.580113986	110576	1921834
Green Junglefowl	*G. g. murghi*	Henan, China	White Plymouth	0.0147	9.69911206	108809	1924411
Green Junglefowl	*G. g. murghi*	Indonesia	White Plymouth	0.0192	14.46711801	113573	1924522
Green Junglefowl	*G. g. murghi*	Iran	White Plymouth	-0.0029	-2.194109461	110091	1924522
Green Junglefowl	*G. g. murghi*	Jiangsu, China	White Plymouth	0.0108	7.357110407	108046	1924410
Green Junglefowl	*G. g. murghi*	Jiangxi, China	White Plymouth	0.0134	10.06911258	109592	1924522
Green Junglefowl	*G. g. murghi*	Liaoning, China	White Plymouth	0.0122	8.353110122	107465	1924508
Green Junglefowl	*G. g. murghi*	Pakistan	White Plymouth	-0.0006	-0.449112999	113136	1924522
Green Junglefowl	*G. g. murghi*	Rhode Island Red	White Plymouth	0.0018	1.077106629	106246	1906711
Green Junglefowl	*G. g. murghi*	Shandong, China	White Plymouth	0.0143	10.7391121	108950	1924522
Green Junglefowl	*G. g. murghi*	Shanxi, China	White Plymouth	0.0064	4.555108964	107573	1924470
Green Junglefowl	*G. g. murghi*	Sri Lanka	White Plymouth	0.0055	3.217112071	110846	1924508
Green Junglefowl	*G. g. murghi*	Taiwan, China	White Plymouth	0.0023	1.275107876	107373	1923779
Green Junglefowl	*G. g. murghi*	Thailand	White Plymouth	0.0215	14.92611965	114610	1924460
Green Junglefowl	*G. g. murghi*	ThaiRJF	White Plymouth	0.0211	15.21312061	115621	1924522
Green Junglefowl	*G. g. murghi*	Thailand-2	White Plymouth	0.0279	19.54012299	116306	1922782
Green Junglefowl	*G. g. murghi*	Tibet, China	White Plymouth	0.0053	4.180116911	115676	1924522
Green Junglefowl	*G. g. murghi*	Vietnam	White Plymouth	0.0217	14.91911936	114278	1924513
Green Junglefowl	*G. g. murghi*	Xinjiang, China	White Plymouth	0.006	4.783110836	109513	1924522
Green Junglefowl	*G. g. murghi*	Yuanbao, China	White Plymouth	0.0118	8.150112531	109895	1924522
Green Junglefowl	*G. g. murghi*	Yunnan, China	White Plymouth	0.0165	12.63911764	113816	1924522
Green Junglefowl	*G. g. murghi*	*G. g. spadiceus*	White Plymouth	0.0175	13.34311995	115823	1924522

Supplementary information, Table S10:
Functional enrichment term for genes with signature of selection by π-ratio.

P-value	Gene number	Term ID	Terms	Description
0.0307	7	GO:0001704	BP	formation of primary germ layer
0.0383	3	GO:0007501	BP	mesodermal cell fate specification
0.05	146	GO:0005575	CC	Cellular component
0.00336	52	HP:0000118	hp	Phenotypic abnormality
0.00258	50	HP:0000005	hp	Mode of inheritance
0.0273	40	HP:0000707	hp	Abnormality of the nervous system
0.0257	38	HP:0012638	hp	Abnormality of nervous system physiology
0.0435	26	HP:0040064	hp	Abnormality of limbs
0.0331	16	HP:0000174	hp	Abnormality of the palate
0.0273	12	HP:0000218	hp	High palate
0.0059	7	HP:0000415	hp	Abnormality of the choanae
0.00437	7	HP:0000453	hp	Choanal atresia
0.00233	5	HP:0001335	hp	Bimanual synkinesia
0.0437	4	HP:0000467	hp	Neck muscle weakness
0.0399	4	HP:0009888	hp	Abnormality of secondary sexual hair
0.00336	4	HP:0100133	hp	Abnormality of the pubic hair
0.00336	4	HP:0002225	hp	Sparse pubic hair
0.0216	4	HP:0100134	hp	Abnormality of the axillary hair
0.0193	4	HP:0002215	hp	Sparse axillary hair
0.0242	4	HP:0030338	hp	Abnormal circulating gonadotropin level
0.00226	4	HP:0030346	hp	Abnormal circulating follicle-stimulating hormone level
0.00277	4	HP:0030339	hp	Decreased circulating gonadotropin level
0.00226	4	HP:0030341	hp	Decreased circulating follicle stimulating hormone level
0.00277	4	HP:0030345	hp	Abnormal circulating luteinizing hormone level
0.00226	4	HP:0030344	hp	Decreased circulating luteinizing hormone level
0.00226	4	HP:0003295	hp	Impaired FSH and LH secretion
0.0298	4	HP:0100299	hp	Muscle fiber inclusion bodies
0.0399	4	HP:0000122	hp	Unilateral renal agenesis
0.00226	4	HP:0002929	hp	Leydig cell insensitivity to gonadotropin
0.00277	4	HP:0003782	hp	Eunuchoid habitus
0.0399	4	HP:0010662	hp	Abnormality of the diencephalon
0.00277	4	HP:0012285	hp	Abnormal hypothalamus physiology
0.00226	4	HP:0003164	hp	Hypothalamic gonadotropin-releasing hormone (GNRH) deficiency
0.00277	4	HP:0012504	hp	Abnormal size of pituitary gland
0.00277	4	HP:0012506	hp	Small pituitary gland
0.0478	4	HP:0001341	hp	Olfactory lobe agenesis
0.0152	4	HP:0030260	hp	Microphallus
0.00277	4	HP:0008197	hp	Absence of pubertal development
0.0193	4	HP:0010311	hp	Aplasia/Hypoplasia of the breasts
0.05	2	HP:0001922	hp	Vacuolated lymphocytes

0.0492	5	KEGG:04261	ke	Adrenergic signaling in cardiomyocytes
0.0455	5	KEGG:04530	ke	Tight junction
0.0006	140	GO:0003674	MF	Molecular function

Supplementary information, Table S11:

Functional enrichment term for genes with signature of selection by LSBL (chicken; *G. g. spadiceus*, *G. g. jabouillei*).

P-value	Gene number	Term ID	Terms	Discription
0.0105	12	GO:0048638	BP	regulation of developmental growth
0.00292	21	GO:0007267	BP	cell-cell signaling
0.0456	2	GO:0098907	BP	regulation of SA node cell action potential
0.0488	4	GO:0051349	BP	positive regulation of lyase activity
0.0488	4	GO:0031281	BP	positive regulation of cyclase activity
0.0488	4	GO:0045761	BP	regulation of adenylate cyclase activity
0.00726	4	GO:0045762	BP	positive regulation of adenylate cyclase activity
0.0135	13	GO:0003013	BP	circulatory system process
0.013	9	GO:0003015	BP	heart process
0.05	12	GO:0008015	BP	blood circulation
0.0153	40	GO:0051239	BP	regulation of multicellular organismal process
0.00278	14	GO:0044057	BP	regulation of system process
0.0222	9	GO:1903522	BP	regulation of blood circulation
0.000286	10	GO:0090257	BP	regulation of muscle system process
0.00204	8	GO:0006937	BP	regulation of muscle contraction
0.05	9	HP:0000453	hp	Choanal atresia
0.05	7	KEGG:04020	ke	Calcium signaling pathway
0.0484	6	KEGG:04261	ke	Adrenergic signaling in cardiomyocytes

Additional references

1. Wang, M.S. *et al.* Genomic Analyses Reveal Potential Independent Adaptation to High Altitude in Tibetan Chickens. *Mol Biol Evol* **32**, 1880-9 (2015).
2. Wang, M.S. *et al.* Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. *Cell Res* **26**, 556-73 (2016).
3. Yi, G. *et al.* Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics* **15**, 962 (2014).
4. Fan, W.L. *et al.* Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol Evol* **5**, 1376-92 (2013).
5. Wang, M.S. *et al.* Comparative population genomics reveals genetic basis underlying body size of domestic chickens. *J Mol Cell Biol* **8**, 542-552 (2016).
6. Wang, M.S. *et al.* An Evolutionary Genomic Perspective on the Breeding of Dwarf Chickens. *Mol Biol Evol* **34**, 3081-3088 (2017).
7. Huang, X.-H. *et al.* Was chicken domesticated in northern China? New evidence from mitochondrial genomes. *Science Bulletin* **63**, 743-746 (2018).
8. Miao, Y.W. *et al.* Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity (Edinb)* **110**, 277-82 (2013).
9. Kong, Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **98**, 152-3 (2011).
10. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843-51 (2014).
11. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
12. Rubin, C.J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587-91 (2010).
13. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
14. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641-50 (2009).
15. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
16. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
17. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
18. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-64 (2009).
19. Brucato, N. *et al.* Evidence of Austronesian Genetic Lineages in East Africa and South Arabia: Complex Dispersal from Madagascar and Southeast Asia. *Genome Biol Evol* **11**, 748-758 (2019).
20. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* **15**, 1179-91 (2015).
21. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
22. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-93 (2012).
23. Reich, D., Thangaraj, K., Patterson, N., Price, A.L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-94 (2009).
24. Pickrell, J.K. & Pritchard, J.K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967 (2012).
25. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**, 919-25 (2014).
26. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-6 (2011).
27. Browning, B.L. & Browning, S.R. Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet* **98**, 116-26 (2016).
28. Malaspinas, A.S. *et al.* A genomic history of Aboriginal Australia. *Nature* **538**, 207-214 (2016).

29. Pagani, L. *et al.* Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am J Hum Genet* **96**, 986-91 (2015).
30. Nam, K. *et al.* Molecular evolution of genes in avian genomes. *Genome Biol* **11**, R68 (2010).
31. Der Sarkissian, C. *et al.* Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr Biol* **25**, 2577-83 (2015).
32. Brisbin, A. *et al.* PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* **84**, 343-64 (2012).
33. Axelsson, E., Webster, M.T., Smith, N.G., Burt, D.W. & Ellegren, H. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res* **15**, 120-5 (2005).
34. Sawai, H. *et al.* The origin and genetic variation of domestic chickens with special reference to junglefowls Gallus g. gallus and G. varius. *PLoS One* **5**, e10639 (2010).
35. Frantz, L.A. *et al.* Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**, 1228-31 (2016).
36. Nishibori, M., Shimogiri, T., Hayashi, T. & Yasue, H. Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius. *Anim Genet* **36**, 367-75 (2005).
37. Peng, M.S. *et al.* DomeTree: a canonical toolkit for mitochondrial DNA analyses in domesticated animals. *Mol Ecol Resour* **15**, 1238-42 (2015).
38. Skoglund, P. *et al.* Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510-513 (2016).
39. de Manuel, M. *et al.* Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* **354**, 477-481 (2016).
40. Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V.C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**, e1003905 (2013).
41. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5**, e1000695 (2009).
42. Bosse, M. *et al.* Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat Commun* **5**, 4392 (2014).
43. Wu, D.D. *et al.* Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nat Ecol Evol* **2**, 1139-1145 (2018).
44. Pei, Y.F., Li, J., Zhang, L., Papasian, C.J. & Deng, H.W. Analyses and comparison of accuracy of different genotype imputation methods. *PLoS One* **3**, e3551 (2008).
45. Gou, X. *et al.* Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res* **24**, 1308-1315 (2014).
46. Freedman, A.H. *et al.* Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genet* **10**, e1004016 (2014).
47. Zhang, W. *et al.* Hypoxia Adaptations in the Grey Wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. *PLoS Genet* **10**, e1004466 (2014).
48. Li, Y. *et al.* Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. *Genome Biol Evol* **6**, 3115-21 (2014).
49. vonHoldt, B., Fan, Z., Ortega-Del Vecchyo, D. & Wayne, R.K. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ* **5**, e3522 (2017).
50. Miao, B., Wang, Z. & Li, Y. Genomic analysis reveals hypoxia adaptation in the Tibetan Mastiff by introgression of the gray wolf from the Tibetan Plateau. *Mol Biol Evol* **34**, 734-743 (2017).
51. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-4 (2016).
52. Shriver, M.D. *et al.* The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics* **1**, 274-86 (2004).
53. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75-8 (2010).
54. Weir, B.S. & Cockerham, C.C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370 (1984).
55. Nei, M. & Li, W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* **76**, 5269-73 (1979).
56. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

57. Ewing, G. & Hermisson, J. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **26**, 2064-5 (2010).
58. Reimand, J., Arak, T. & Vilo, J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* **39**, W307-15 (2011).
59. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-52 (2015).
60. Naval-Sanchez, M. *et al.* Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat Commun* **9**, 859 (2018).
61. Carneiro, M. *et al.* Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074-9 (2014).
62. Frantz, L.A. *et al.* Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nature Genetics* **47**, 1141-1148 (2015).
63. Axelsson, E. *et al.* The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360-4 (2013).
64. Wang, X. *et al.* Genomic responses to selection for tame/aggressive behaviors in the silver fox (Vulpes vulpes). *Proc Natl Acad Sci U S A* **115**, 10398-10403 (2018).
65. Montague, M.J. *et al.* Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 17230-17235 (2014).
66. Sutter, N.B. *et al.* A single IGF1 allele is a major determinant of small size in dogs. *Science* **316**, 112-5 (2007).
67. Sallinen, J., Haapalinna, A., Viitamaa, T., Kobilka, B.K. & Scheinin, M. Adrenergic α_{2C}-Receptors Modulate the Acoustic Startle Reflex, Prepulse Inhibition, and Aggression in Mice. *The Journal of Neuroscience* **18**, 3035 (1998).
68. Ninan, I. & Arancio, O. Presynaptic CaMKII is necessary for synaptic plasticity in cultured hippocampal neurons. *Neuron* **42**, 129-41 (2004).
69. Qanbari, S. *et al.* Genetics of adaptation in modern chicken. *PLoS Genet* **15**, e1007989 (2019).
70. Darwin, C. The variation of animals and plants under domestication *London, UK: John Murray* (1868).
71. West, B. & Zhou, B.-X. Did chickens go North? New evidence for domestication. *Journal of Archaeological Science* **15**, 515-533 (1988).
72. Lawler, A. Animal domestication. In search of the wild chicken. *Science* **338**, 1020-4 (2012).
73. Nishida, T. *et al.* Morphological Identification and Ecology of the Red Jungle Fowl in Thailand, Laos and Vietnam. *Nihon Chikusan Gakkaiho* **71**, 470-480 (2000).
74. Akram, S., Awan, M.S., Maqsood, A., Usman, A. & Dar, N.I. Population status and distribution pattern of Red Jungle Fowl (Gallus gallus murghi) in Deva Vatala National Park, Azad Jammu & Kashmir, Pakistan: a pioneer study. *Pakistan Journal of Zoology* **42**, 701-706 (2010).
75. Tixier-Boichard, M., Bed'hom, B. & Rognon, X. Chicken domestication: from archeology to genomics. *C R Biol* **334**, 197-204 (2011).
76. Fumihito, A. *et al.* Monophyletic origin and unique dispersal patterns of domestic fowls. *Proc Natl Acad Sci U S A* **93**, 6792-5 (1996).
77. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
78. Wang, K., Mathieson, I., O'Connell, J. & Schiffels, S. Tracking human population structure through time from whole genome sequences. *bioRxiv*, 585265 (2019).
79. Hwang, J.S. The genes associated with gonadotropin-releasing hormone-dependent precocious puberty. *Korean J Pediatr* **55**, 6-10 (2012).
80. Lawal, R.A. *et al.* Whole-Genome Resequencing of Red Junglefowl and Indigenous Village Chicken Reveal New Insights on the Genome Dynamics of the Species. *Front Genet* **9**, 264 (2018).