UC Santa Cruz
UC Santa Cruz Previously Published Works

Title
863 genomes reveal the origin and domestication of chicken

Permalink
https://escholarship.org/uc/item/6gs8n86k

Journal
Cell Research, 30(8)

ISSN
1001-0602

Authors
Wang, Ming-Shan
Thakur, Mukesh
Peng, Min-Sheng
et al.

Publication Date
2020-08-01

DOI
10.1038/s41422-020-0349-y

Peer reviewed

ARTICLE

863 genomes reveal the origin and domestication of chicken

Ming-Shan Wang[1,2,36], Mukesh Thakur [1,3], Min-Sheng Peng [1,2,4], Yu Jiang[5], Laurent Alain François Frantz [1,6,7], Ming Li [5], Jin-Jin Zhang[1,2], Sheng Wang[1,2], Joris Peters[8,9], Newton Otieno Otecko[1,2], Chatmongkon Suwannapoom[10], Xing Guo[11], Zhu-Qing Zheng[5], Ali Esmailizadeh [1,12], Nalini Yasoda Hirimuthugoda[1,13], Hidayat Ashari[14,15], Sri Suladari[14], Moch Syamsul Arifin Zein[14], Szilvia Kusza[16], Saeed Sohrabi[12], Hamed Kharrati-Koopaee[12,17], Quan-Kuan Shen [1,2], Lin Zeng[1,2], Min-Min Yang[1,2], Ya-Jiang Wu[1,18], Xing-Yan Yang[1,18], Xue-Mei Lu[1,2,4], Xin-Zheng Jia[19,20], Qing-Hua Nie[21], Susan Joy Lamont[20], Emiliano Lasagna[22], Simone Ceccobelli[22], Humpita Gamaralalage Thilini Nisanka Gunwardana[13], Thilina Madusanka Senasige[13], Shao-Hong Feng[23], Jing-Fang Si[24], Hao Zhang[24], Jie-Qiong Jin[1,25], Ming-Li Li[1,2], Yan-Hu Liu[1,2], Hong-Man Chen[1,2], Cheng Ma[1,2], Shan-Shan Dai[1,2], Abul Kashem Fazlul Haque Bhuiyan[26], Muhammad Sajjad Khan[27], Gamamada Liyanage Lalanie Pradeepa Silva[28], Thi-Thuy Le[29], Okeyo Ally Mwai[19], Mohamed Nawaz Mohamed Ibrahim[19], Megan Supple[30], Beth Shapiro [30,31], Olivier Hanotte[32,33], Guojie Zhang [1,4,34,35], Greger Larson [6], Jian-Lin Han [15,19], Dong-Dong Wu [1,2,4] and Ya-Ping Zhang [1,2,4,18]

Despite the substantial role that chickens have played in human societies across the world, both the geographic and temporal origins of their domestication remain controversial. To address this issue, we analyzed 863 genomes from a worldwide sampling of chickens and representatives of all four species of wild jungle fowl and each of the five subspecies of red jungle fowl (RJF). Our study suggests that domestic chickens were initially derived from the RJF subspecies *Gallus gallus spadiceus* whose present-day distribution is predominantly in southwestern China, northern Thailand and Myanmar. Following their domestication, chickens were translocated across Southeast and South Asia where they interbred locally with both RJF subspecies and other jungle fowl species. In addition, our results show that the White Leghorn chicken breed possesses a mosaic of divergent ancestries inherited from other subspecies of RJF. Despite the strong episodic gene flow from geographically divergent lineages of jungle fowls, our analyses show that domestic chickens undergo genetic adaptations that underlie their unique behavioral, morphological and reproductive traits. Our study provides novel insights into the evolutionary history of domestic chickens and a valuable resource to facilitate ongoing genetic and functional investigations of the world's most numerous domestic animal.

INTRODUCTION

Since their domestication, chickens (*Gallus gallus domesticus*) have been venerated by diverse cultures across the world. Relative to other domestic animals including sheep, cattle and pigs, chickens are currently both the preferred source of animal protein and the most numerous domestic animal.[1] Despite their popularity and

[1]State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; [2]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan, China; [3]Zoological Survey of India, New Alipore, Kolkata, West Bengal, India; [4]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan, China; [5]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, China; [6]The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK; [7]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK; [8]ArchaeoBioCenter and Department of Veterinary Sciences, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, LMU Munich, Munich, Germany; [9]SNSB, Bavarian State Collection of Anthropology and Palaeoanatomy, Munich, Germany; [10]School of Agriculture and Natural Resources, University of Phayao, Phayao, Thailand; [11]College of Animal Science and Technology, Anhui Agricultural University, Hefei, Anhui, China; [12]Department of Animal Science, Shahid Bahonar University of Kerman, Kerman, Iran; [13]Faculty of Agriculture, University of Ruhuna, Matara, Sri Lanka; [14]Museum Zoologicum Bogoriense, Research Center for Biology, Indonesian Institute of Science (LIPI), Cibinong, Indonesia; [15]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China; [16]Institute of Animal Husbandry, Biotechnology and Nature Conservation, University of Debrecen, Debrecen, Hungary; [17]Institute of Biotechnology, School of Agriculture, Shiraz University, Shiraz, Iran; [18]State Key Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, Yunnan, China; [19]Livestock Genetics Program, International Livestock Research Institute (ILRI), Nairobi, Kenya; [20]Department of Animal Science, Iowa State University, Ames, IA, USA; [21]College of Animal Science, South China Agricultural University, Guangzhou, Guangdong, China; [22]Dipartimento di Scienze Agrarie, Alimentare Ambientali, University of Perugia, Perugia, Italy; [23]BGI-Shenzhen, Beishan Industrial Zone, Shenzhen, Guangdong, China; [24]Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture of China, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China; [25]Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences (CAS-SEABRI), Yezin, Myanmar; [26]Bangladesh Agricultural University, Mymennsingh, Bangladesh; [27]University of Agriculture Faisalabad, Faisalabad, Pakistan; [28]Department of Animal Science, University of Peradeniya, Peradeniya, Sri Lanka; [29]National Institute of Animal Husbandry, Hanoi, Vietnam; [30]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA, USA; [31]Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, CA, USA; [32]Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, UK; [33]Livestock Genetics Program, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia; [34]Department of Biology, Centre for Social Evolution, University of Copenhagen, Copenhagen, Denmark and [35]China National Genebank, BGI-Shenzhen, Shenzhen, Guangdong, China
Correspondence: Jian-Lin Han (h.jianlin@cgiar.org) or Dong-Dong Wu (wudongdong@mail.kiz.ac.cn) or Ya-Ping Zhang (zhangyp@mail.kiz.ac.cn)
[36]Present address: Department of Ecology and Evolutionary Biology, Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, CA, USA
These authors contributed equally: Ming-Shan Wang, Mukesh Thakur, Min-Sheng Peng, Yu Jiang, Laurent Alain François Frantz, Ming Li

ubiquity, both the geographic and temporal origins of domestic chickens remain controversial. The red jungle fowl (RJF, *G. gallus*; Supplementary information, Fig. S1) is believed to be the wild progenitor of domestic chickens, and chicken domestication is thought to have occurred during the Holocene.[2–4] Which subspecies of extant RJF (*G. g. gallus*, *G. g. spadiceus*, *G. g. jabouillei*, *G. g. murghi*, and *G. g. bankiva*) was first domesticated and to what degree domestic chickens interbred with other (sub)species of jungle fowl remain unresolved questions.[4–8]

From an archaeological perspective, a significant challenge has been how to confidently identify chickens since no osteomorphological markers can readily distinguish the five RJF subspecies from each other, or discriminate between RJF and early domestic chickens.[3,9,10] Additionally, attempts to characterize the spatio-temporal origins and subsequent dispersals of chickens have been hampered by a lack of direct radiocarbon dating of presumed early archaeological remains.[9,11]

Numerous genetic studies based on mitochondrial DNA (mtDNA) sequences raised the "multiple origins" hypothesis, which claimed that wild RJFs were incorporated into ancient food-producing cultures in multiple occasions.[5,6] However, the general propensity of domestic animals to admix with their wild relatives, including those that were never independently domesticated, can lead to spurious claims of multiple and independent origins based on this single genetic marker.[2,12] Additionally, as a maternally inherited non-recombining DNA, mtDNA has a limited power to reveal complex past demography.[13] Conversely, the rapid development of whole-genome sequencing holds a great promise for inferring the evolutionary history of domestication processes.[14–19]

Over the past decade, several population genomic studies have been conducted to investigate the genetic basis underlying the process of chicken domestication. However, most studies have largely focused on either commercial breeds[20–22] or specific local populations,[23–26] and most have been performed with limited genomic data from RJF. The genetic divergence and structure of different RJF subspecies has thus far not been conducted. Without access to these datasets, previous studies have had limited power to infer the spatiotemporal origins and genetic adaptions underlying the domestication of domestic chickens.

To establish the primary RJF subspecies from which domestic chickens were derived (and hence infer their geographic origin), and to understand the genetic mechanisms underlying chicken domestication, it is necessary to analyze the nuclear genomes of both presumed wild relatives and domestic populations, within and beyond the natural distribution ranges of all RJF subspecies. Here, we inferred the history of chicken domestication and the genetic signatures of selection in domestic chicken through a large-scale whole-genome sequencing of domestic chickens collected from a global sampling and all of the wild jungle fowl species.

RESULTS

Phylogeny and admixture between jungle fowl species and RJF subspecies

We sequenced 787 whole genomes: 627 domestic chickens, 142 RJFs representing all five subspecies, 12 green jungle fowls (*G. varius*), 2 gray jungle fowls (*G. sonneratii*) and 4 Ceylon jungle fowls (*G. lafayettii*) (Fig. 1a; Supplementary information, Table S1). To maximize the likelihood of capturing genetic variability among RJF subspecies, we sampled individuals belonging to each subspecies from at least three geographically distant locations and ensured that at least one individual of each subspecies was sequenced to at least 20× coverage (Supplementary information, Table S1). While it was not possible to ensure that any RJF lineage was completely un-admixed with domestic chickens, this extensive sampling and sequencing strategy should mitigate the

potentially confounding effects of recent admixture on the determination of the origin and early domestication history of chickens. By analyzing these genomes in combination with 76 previously published genomes[24,26–29] (including seven RJFs and 69 chickens), we discovered more than 33.4 million bi-allelic SNPs representing the most comprehensive catalog of genetic diversity for domestic chickens and wild jungle fowls to date, and we deposited genomic data into the ChickenSD database (http://bigd. big.ac.cn/chickensd/). Our dataset includes almost all previously identified RJF and chicken mtDNA haplogroups[5,30] as well as unclassified lineages (Supplementary information, Table S1 and Fig. S2), suggesting that these samples represent modern genetic diversity of wild RJFs and domestic chickens.

First, we clarified the evolutionary history of RJFs, so as to determine whether recent admixture among the five subspecies might obfuscate the timing and location of chicken domestication. A summarized phylogeny of 149 RJF nuclear genomes clusters nearly all of these genomes (with the exception of the *G. g. murghi* lineage containing three out of 27 *G. g. jabouillei* birds sampled from Guangxi province of China) into five discrete clades (Fig. 1b; Supplementary information, Figs. S3 and S4). In this analysis, *G. g. bankiva*, sampled in eastern Java, is basal to all RJF subspecies. Principal component analysis (PCA) also highlights a separation among RJF subspecies (Supplementary information, Figs. S5-S9). It is interesting that some *G. g. murghi* (distributed across the northern Indian subcontinent) and *G. g. jabouillei* (confined to South China and North Vietnam) individuals cluster together, since their present-day distribution ranges are separated by *G. g. spadiceus* whose present range covers predominately south-western China, northern Thailand and Myanmar. Pairwise F_{ST} estimates also show that *G. g. murghi* is more closely related to *G. g. jabouillei* than to *G. g. spadiceus* (Supplementary information, Fig. S10).

We then used ADMIXTURE[31] to identify the patterns of genetic clustering and found that some *G. g. murghi* individuals share more ancestral components with *G. g. jabouillei*, though this analysis and PCA are possibly confounded by the numbers of samples and populations used in the analyses (Supplementary information, Figs. S11-S13). *D*-statistic analyses further suggest a complex genetic relationship among *G. g. spadiceus*, *G. g. jabouillei* and *G. g. murghi* (Supplementary information, Fig. S14). In addition, the results from TreeMix[32] and qpGraph[33] revealed that RJF subspecies show evidence of admixture, e.g., between *G. g. spadiceus* and *G. g. gallus*, and between *G. g. gallus* and *G. g. bankiva* (Supplementary information, Figs. S15, S16 and Table S2), indicating a long history of gene flow between RJF subspecies.

Overall, our analyses indicate that all RJF subspecies are genetically differentiated, which generally correspond to their geographic ranges and taxonomic classifications. In order to assess the timing of the divergence between the different RJF subspecies, we used multiple sequential Markovian coalescent (MSMC),[34] assuming a generation time (g) of one year and mutation rate (μ) of 1.91×10^{-9} per generation.[35] We restricted these analyses to samples with > 20× sequencing coverage. This analysis (based on a 50% relative cross coalescence rate (CCR) cutoff) indicates that *G. g. bankiva* is the most divergent subspecies and has a time to the most recent common ancestor (TMRCA) with the other RJF subspecies prior to 500 kya (Supplementary information, Fig. S17a), consistent with its basal phylogenetic position. The TMRCA of the four other RJF subspecies was between 50 and 125 kya (Supplementary information, Fig. S17b). These analyses indicate that all of the RJF subspecies diverged from one another substantially earlier than the advent of chicken domestication.[2–4]

Geographic origin of domestic chickens

We then sought to identify the specific RJF lineage(s) from which domestic chickens were derived. The phylogeny constructed with
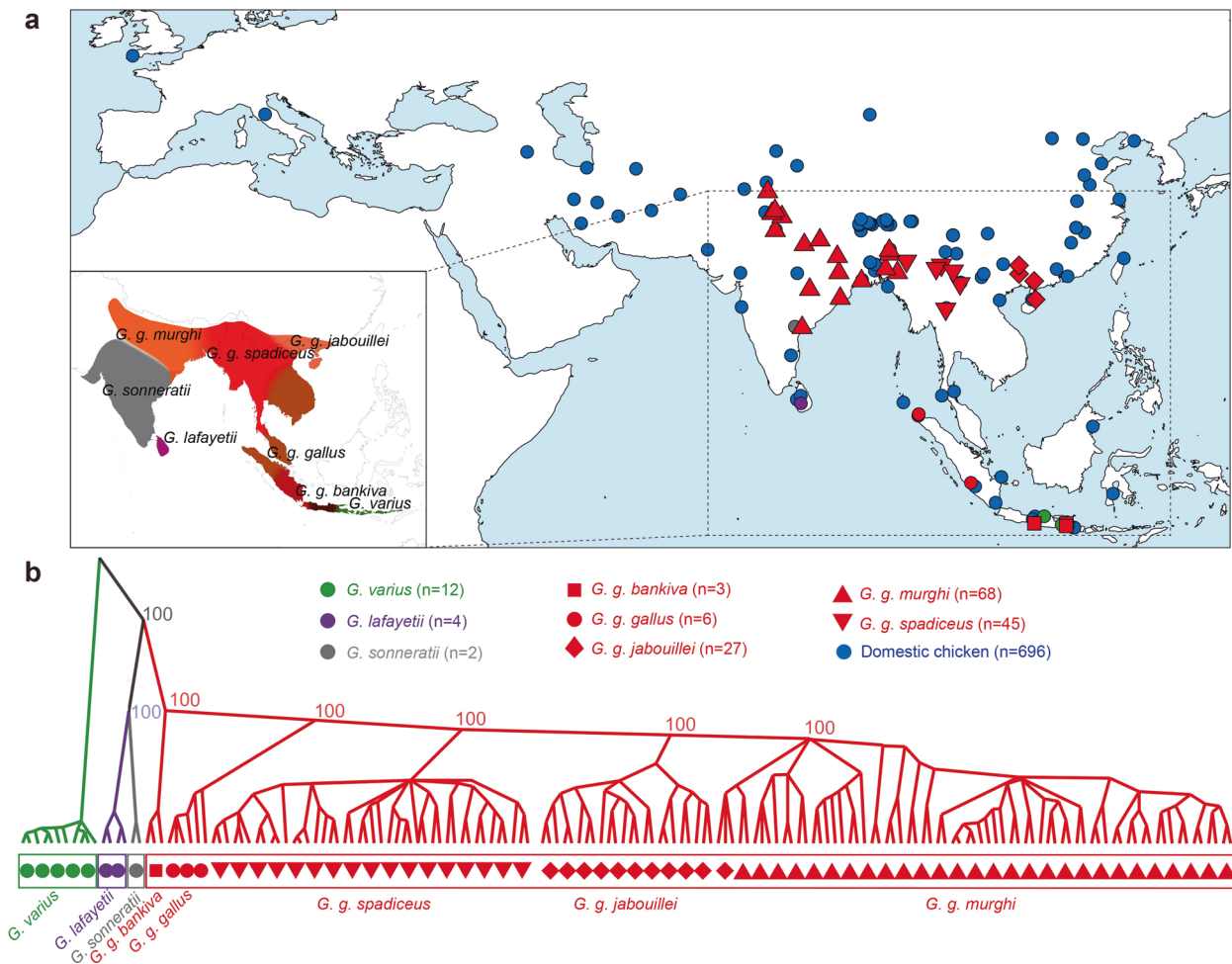
Fig. 1 **Sample distribution and phylogeny of *Gallus* taxa. a** Map showing the geographic distribution of the sampling localities of all domestic chickens and wild jungle fowls across South Asia, Southeast Asia, and East Asia covered by our study. Map in the left bottom depicts the natural ranges of all jungle fowls (see details in Supplementary information, Fig. S1). **b** Maximum-likelihood tree depicting the evolutionary relationships among RJF subspecies and other three jungle fowl species. The numbers at the major branches are bootstrap values.

all 149 RJFs and 696 domestic chickens supports a monophyletic clade composed of some wild *G. g. spadiceus* specimens and all but two of the 696 domestic chickens (Fig. 2a). Interestingly, the wild *G. g. spadiceus* that fall within, and the two chickens that fall outside this clade were sampled in Thailand (Supplementary information, Figs. S18 and S19). f_4-statistics indicate that these exceptions are the result of gene flow between wild RJFs and domestic chicken populations (Supplementary information, Table S3). This finding is consistent with observations that domestic village chickens were hybridized with wild *G. g. spadiceus* in Thailand in the mid-20th century, and wild RJF clutches were removed from their nests and hatched by domestic hens.[36]

Of the five RJF subspecies, individuals of *G. g. spadiceus* are the most closely related to all domestic chicken populations (Fig. 2a; Supplementary information, Fig. S20). Further, PCA, ADMIXTURE, as well as outgroup-*f3* and *f4* analyses (Figs. 2b, 3a; Supplementary information, Figs. S21-S29) also unequivocally indicate that domestic chickens cluster more closely with *G. g. spadiceus* than with the other four RJF subspecies. Finally, MSMC analysis indicates that the split of *G. g. spadiceus* from domestic chickens took place ~9500 ± 3300 years ago (Fig. 3b, c; Supplementary information, Figs. S30 and S31). By combining the monophyletic nature of all domestic chickens, the results from these analyses collectively suggest that chickens were

likely domesticated in the Holocene from the *G. g. spadiceus* subspecies of RJF.

We also identified two well-defined clades: I and II (Fig. 2a; Supplementary information, Fig. S32). Clade I includes chickens from Europe and the Americas (including European broiler and egg layer chickens of White Leghorn, White Plymouth, Rhode Island Red and Cornish breeds), Iran, Pakistan, India, Bangladesh and northwestern China (i.e., Tibet and Xinjiang provinces bordering India). Clade II contains mostly northern, central, and southern Chinese village chickens (i.e., from Shanxi and Jiangxi provinces). Branches basal to the two clades, but within the total diversity of chickens, include 128 chickens sampled almost exclusively from the Yunnan province of China, Thailand, Vietnam and Indonesia. These individuals may represent the earliest domestic lineages or have admixed with local RJF subspecies.

Dispersal and admixture patterns of domestic chickens
Our results contradict previous claims that chickens were domesticated in Neolithic northern China[37] and the Indus Valley Civilization (made on the basis of suspected chicken remains found at the site of Mohenjo-Daro in Pakistan).[38] However, a PCA shows that *G. g. murghi* samples from westernmost North India showed a deeper divergence from chickens than the remaining birds of *G. g. murghi* collected from northeastern India (Fig. 2b). Moreover, our mtDNA analyses revealed that the most frequent
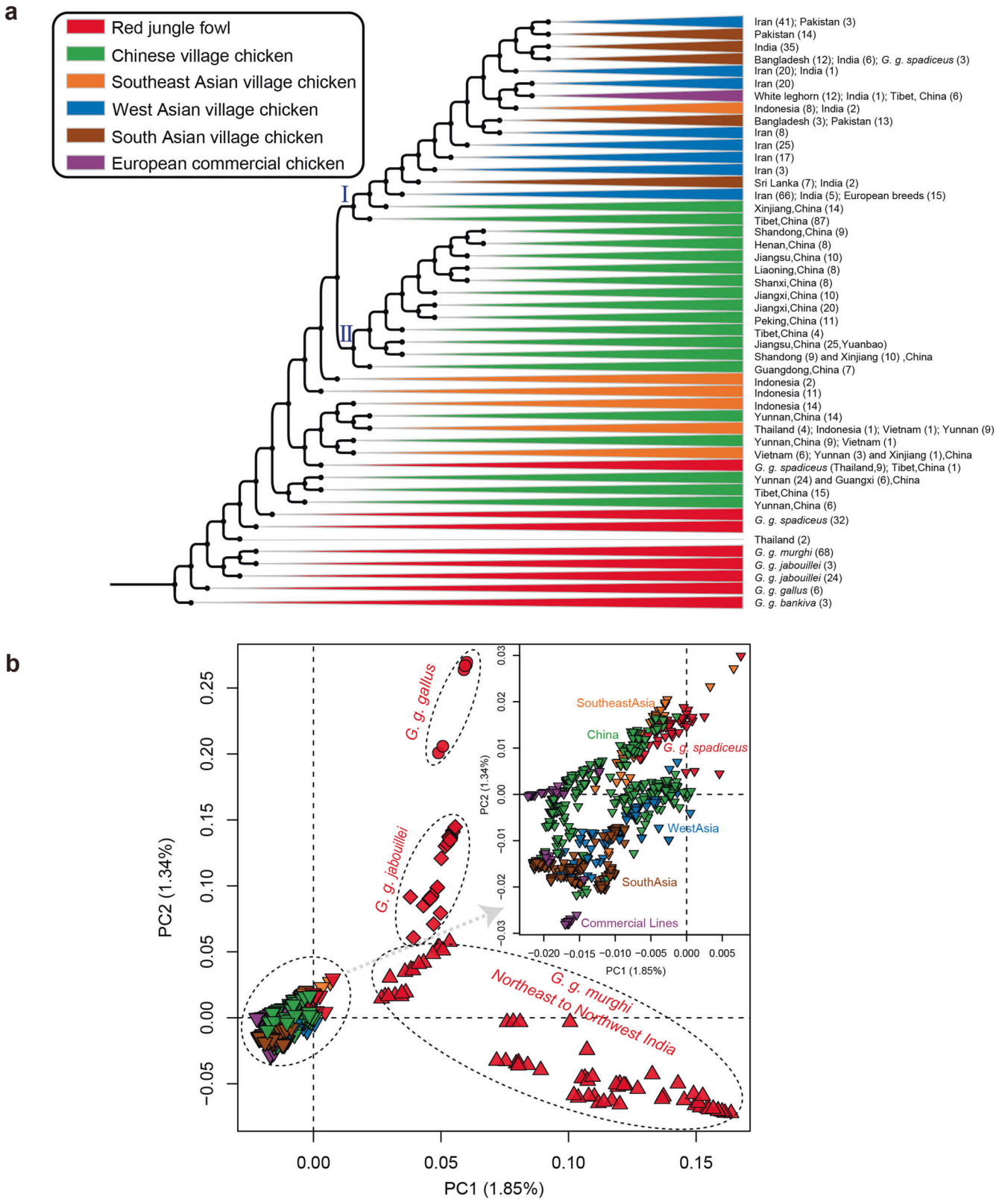
Fig. 2 Domestic chickens were most likely derived from *G. g. spadiceus*. **a** Maximum-likelihood phylogenetic tree showing that domestic chickens form a monophyletic clade, with *G. g. spadiceus* being the closest wild progenitor. Black dots at nodes indicate ≥ 99% bootstrap support. Domestic chicken and RJF clades are collapsed and colored according to their geographic ranges and subspecies classifications. **b** PCA showing a closer genetic affinity between domestic chickens and *G. g. spadiceus*. *G. g. bankiva* was removed in this analysis because of its high divergence from other four RJF subspecies. RJF subspecies are denoted within rings.
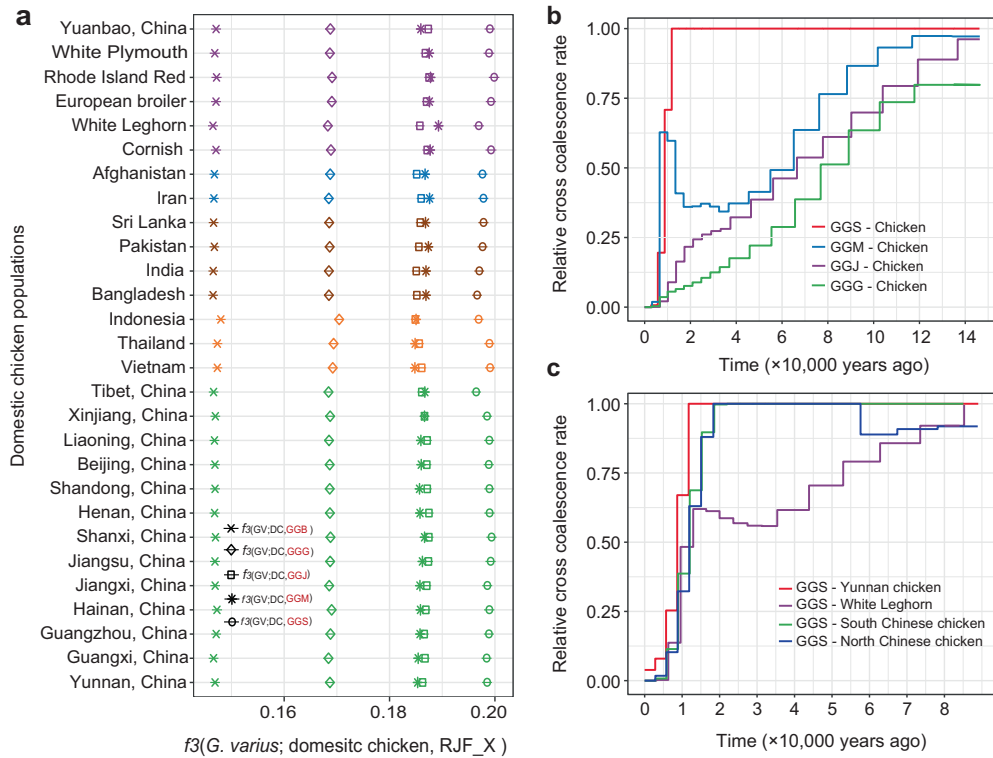
Fig. 3 The admixture and splitting of domestic chickens with RJF subspecies. a Outgroup-*f*3 statistics in the form of *f*3(*G. varius*; domestic chicken, RJF) show that all domestic chickens carry more genetic ancestry from *G. g. spadiceus* (GGS, higher *f*3 values) than from other four RJF subspecies. The estimated *f*3 value ± 3 standard errors are plotted. **b** MSMC plots show the divergence time between chicken and each of RJF subspecies. **c** MSMC plots show the splitting time between chicken and *G. g. spadiceus*. For clarity, we only present one result of MSMC for each population pair; more pairs were analyzed and shown in Supplementary information, Figs. S30, S31 and S48a. GGB *G. g. bankiva*, GGG *G. g. gallus*, GGM *G. g. murghi*, GGJ *G. g. jabouillei*.

and dominant haplogroups of South Asian chickens are D and E, which are similar to that from Southeast Asia and China, but seldom detected in *G. g. murghi* (Supplementary information, Fig. S2 and Table S1).

MSMC estimate indicates that the divergence time between domestic chickens and *G. g. murghi* is ~54.8 ± 5.1 kya (Fig. 3b; Supplementary information, Fig. S30), similar to that between *G. g. murghi* and *G. g. spadiceus* (Supplementary information, Fig. S17). This deeper timeframe of divergence relative to the split between *G. g. spadiceus* and domestic chickens shows that *G. g. murghi* was not the primary source from which domestic chickens were derived. Because introgression following domestication is common,[13] we therefore assessed the potential contribution of *G. g. murghi* to the gene pool of domestic chickens using outgroup-*f*3 and *f*4 statistics (Fig. 3a; Supplementary information, Figs. S29, S33-S35). These analyses indicate that *G. g. murghi* contributed 3.8%–22.4% of the ancestry of domestic chickens from South Asia, particularly those from India (~17.6%), Pakistan (~8.4%), and Bangladesh (~22.4%) (Supplementary information, Fig. S36).

Taken together, these analyses suggest that *G. g. murghi* was not the primary source of domestic chickens, but that this subspecies made a substantial genetic contribution to domestic chickens via gene flow following their domestication in Southeast Asia. Alternatively, chickens may have been domesticated from *G. g. murghi*, but subsequently replaced by birds descended from *G. g. spadiceus*. To test this possibility, we used PCAdmix[39] to compare the lengths of haplotype blocks in the genomes of Indian chickens that are shared with both *G. g. spadiceus* and *G. g. murghi* (Supplementary information, Fig. S37). We observed that Indian chickens share significantly smaller haplotype blocks with *G. g. spadiceus* than with *G. g. murghi* (*P* < 2.2e-16 by Student's *t*-test), a pattern more readily explained by gene flow from *G. g. murghi* to

Indian chickens following their primary origin from *G. g. spadiceus*. In addition, qpGraph, TreeMix and fastsimicoal2[40] analyses also favor a model in which all domestic chickens were initially derived from *G. g. spadiceus* but not from *G. g. murghi* or other RJF subspecies (Supplementary information, Figs. S38-S43). Thus, we show that chickens were unlikely initially domesticated from *G. g. murghi* and subsequently replaced by birds descended from *G. g. spadiceus*.

Multiple lines of analyses, including outgroup-*f*3 and *f*4 statistics, TreeMix and qpGraph, indicate that admixture between RJF subspecies and domestic chickens is common. For example, Indonesian chickens inherit 1.6%–6.5% ancestry from *G. g. bankiva* and 4.8%–10.7% ancestry from *G. g. gallus*, while Chinese chickens possess 1.3%–6.2% ancestry from *G. g. jabouillei* (Supplementary information, Fig. S36). These admixture signals, however, do not always match expectations based solely on the geographic distributions of each of the RJF subspecies. For example, commercial White Leghorns sampled in Iran, China, Indonesia, the United States and Italy (Supplementary information, Table S4) derive ~25% ancestry from *G. g. murghi*, a proportion that is significantly higher than that in any other domestic chicken population, including South Asian ones (Fig. 3a; Supplementary information, Figs. S36, S44-S46 and Tables S5-S9). A PCAdmix analysis revealed that the lengths of haplotype blocks shared between White Leghorn and *G. g. murghi* are significantly larger than those shared between White Leghorn and *G. g. spadiceus* (Supplementary information, Fig. S47; *P* < 2.2e-16 by Student's *t*-test), suggesting that the ancestry from *G. g. murghi* derives from a recent introgression, a pattern also supported by MSMC analysis (Supplementary information, Fig. S48). All these analyses demonstrate that the contribution from *G. g. murghi* played a key role in the development of the White Leghorn.
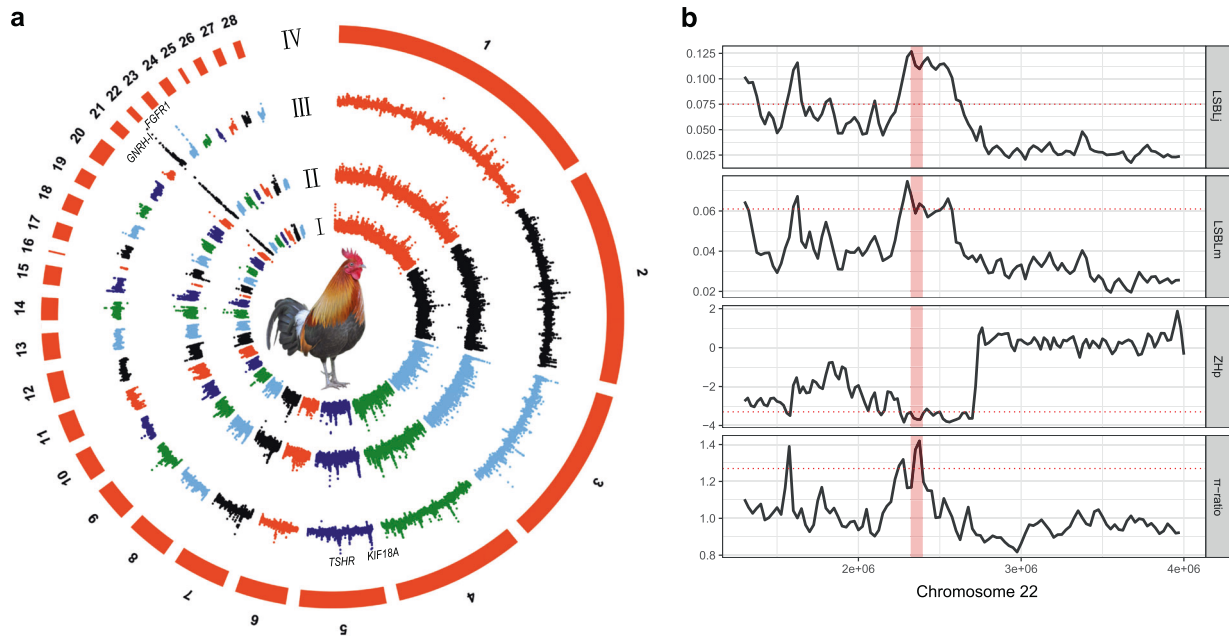
Fig. 4 Signatures of selection in domestic chickens. a Genomic landscape of selection signal in domestic chicken. From inner to outer, circle indicates signature of selection from each statistics: π-ratio (I), LSBL$_j$(chicken; *G. g. spadiceus*, *G. g. jabouillei*) (II), LSBL$_m$(chicken; *G. g. spadiceus*, *G. g. murghi*) (III), and chromosome scheme (IV), respectively. **b** Signals of positive selection on *FRGR1*. *FRGR1*-located region in chicken genome showed a lower diversity (π-ratio), lower heterozygosity (Hp) and higher differentiation (LSBL) compared with RJFs. Red dashed line indicates Z-transformed score of 3.3 for each statistic, and pink shadow depicts the location of *FGFR1* on the chr22.

Previous studies suggested that three additional species of jungle fowls likely contributed to the genetic make-up of modern domestic chickens.[41,42] To test this hypothesis, we identified shared identity-by-decent (IBD) blocks in the genome of wild and domestic fowls using Beagle (Supplementary information, Fig. S49).[43] Using a cutoff of two standard deviations from the mean of the Z-transformed IBD distribution,[44] we found evidence of admixture between these jungle fowl species and domestic chickens (Supplementary information, Figs. S50–S52). However, these introgressed fragments occur at very low frequency and are primarily limited to local chickens that inhabit the native ranges of the local wild jungle fowl species (e.g., green jungle fowl with Indonesia chicken, Ceylon jungle fowl with Sri Lankan chicken), except for the gray jungle fowl. It is plausible, however, that a portion of these signals is misleading since our gray jungle fowl samples were obtained from a zoo population, which may have been admixed previously with chickens. Overall, consistent with the previous study,[42] our analyses suggest that though other jungle fowl species have contributed to the genetic make-up of some local chicken populations, the admixed genomic proportions are very limited.

Patterns of selection in domestic chickens
We used our extensive dataset to identify genomic regions that were affected by positive selection in domestic chickens. We leveraged the locus-specific branch length (LSBL) statistics[45] and π-ratios.[46] Genes under selection were identified based on Z-transforming score ≥ 3.3 (Fig. 4a; Supplementary information, Fig. S53). Through these analyses, we found that genes bearing signal of selection are associated with development of nervous system, muscle and bone as well as regulation of growth, metabolism and reproduction (see a discussion of these genes in Supplementary information, Notes and Tables S10, S11). Interestingly, multiple genes with evidence of selection are found in the neural crest development pathway, including *FGFR1* (fibroblast growth factor receptor 1), *MYC-l*, *ERBB4*, and *BMPs*. *FGFR1* (Fig. 4b) plays an essential role in the regulation of embryonic development and skeletogenesis, and has also been

shown to be under selection in other domestic animals including horse[47] and carp.[48]

Domestic chickens are generally more fertile, produce more eggs and mature earlier than their wild counterparts.[49–51] Our selection analyses identified several genes that are involved in reproductive processes, including *GNRH-I* (gonadotropin-releasing hormone 1) and *KIF18A* (kinesin family member 18A) (Fig. 4a; Supplementary information, Fig. S54–S57). *GNRH-I* is a principal regulator in the reproductive axis controlling onset of puberty and sexual maturity.[52–54] *KIF18A* is known for its role in controlling mitotic expansion and spermatogonial cell differentiation during testis maturation[55] and in determining reproduction ability.[56] No nonsynonymous mutation was found in these genes, suggesting that selection acted on their regulatory elements.

A missense mutation within the *TSHR* gene (thyroid-stimulating hormone receptor; chr5:40,089,599 G/A: TSHR-Gly558Arg) was previously suspected to be a domestication locus based on a preliminary analysis that showed its near fixation in domestic chickens and its virtual absence in RJF.[21] A subsequent analysis of ancient DNA derived from European archaeological chickens revealed that the frequency of this allele began increasing dramatically about 1000 years ago and only reached fixation recently.[57] Interestingly, this mutation was found at high frequency in *G. g. spadiceus* (94.0%) and in Thai RJF (unclassified subspecies previously reported (90.5%),[58] but only 5.4% in other RJF subspecies (Supplementary information. Fig. S58). In addition, we also identified a 239 bp deletion (chr5:40,080,509–40,080,747) in the 7th intron of *TSHR* that shows a similar frequency pattern to TSHR-Gly558Arg in chickens and RJFs (Supplementary information, Fig. S59), suggesting that the two mutations are likely genetically linked.

DISCUSSION

In this study, we present, to the best of our knowledge, the largest genome sequencing initiative for domestic chickens and all wild jungle fowl (sub)species at a global scale to date. Our analyses suggest that domestic chickens were derived initially from the wild RJF subspecies *G. g. spadiceus* that are currently indigenous in

southwestern China, Thailand and Myanmar. A molecular clock analysis suggests that domestic chickens diverged from *G. g. spadiceus* ~9500 ± 3300 years ago, though this node does not necessarily correlate with the beginning of domestication process, as chickens are archaeologically visible much later.[9] This is similar to modern wolves and dogs, whose divergence time estimated based on whole genomes is ~15,000 years earlier than the accepted evidence of domestic dogs in the archeological record.[14,16]

Curiously, the split time between chickens and *G. g. spadiceus* coincides with a period of major climate shifts both globally, with the transition to the Holocene, and locally, with increased temperatures and monsoon activities in southern China.[59,60] These shifts in climates and available habitats may have led to a diversification within *G. g. spadiceus*, followed by the domestication of its specific lineage(s). Since our sample set does not include representatives from every single extant population, we have yet to establish how many *G. g. spadiceus* lineages were involved in the initial domestication process.

The results of our whole-genome analyses also indicate that the five RJF subspecies form monophyletic clades. Continuous post-divergence gene flow is found for RJF populations, especially for those with overlapping ranges, similar to numerous wild canid populations.[61] Particularly, it is striking that the genetic relationships do not always correlate with the current geographic distributions of the RJF subspecies. For example, modern *G. g. jabouillei* and *G. g. murghi* are geographically separated by *G. g. spadiceus*, but the former two subspecies show a close genetic relationship (Figs. 1, 2). One possible explanation for this pattern is that today's distribution ranges of these three RJF subspecies may be different from those in the past. *G. g. jabouillei* and *G. g. murghi* may have historically overlapped while *G. g. spadiceus* probably occupied a geographic region further south. Past contraction and/or expansion of these RJF subspecies, potentially *G. g. spadiceus*'s expansion to further north, may have allowed for their admixture and rapid differentiation. This scenario may be possible since these RJF subspecies diverged from each other ~50–80 kya (Supplementary information, Fig. S17), which is consistent with the expansion of animals in north equatorial Southeast Asia that was possibly facilitated by climatic fluctuations during the Last Glacial Period (10–125 kya).[62,63]

Following their domestication, chickens were then translocated across Southeast and South Asia where they interbred with highly divergent local RJF subspecies and other jungle fowl species. Domestic chickens in China, Southeast Asia and South Asia now all possess hybrid genomes that derive up to 22.4% of their genetic make-up from RJF subspecies other than *G. g. spadiceus*. For example, the genetic make-up of White Leghorns shows a substantial contribution from *G. g. murghi*, and this signature is further detected in some local improved lineages in China and South Asia (Supplementary information, Figs. S33-S36). Given the evolutionary history of chicken accompanied by episodes of recurrent hybridization, it is not surprising that the mitogenomes of wild relatives and domestic chickens were shared between lineages.[5,6] Previous studies[5–7,64] that relied on this single genetic marker therefore had limited power to detect the origins and routes of dispersal of domestic chickens.

Despite this interbreeding, our analyses identified multiple genes involved in behavior, growth and reproduction in domestic chickens that bear signatures of positive selection. Previous studies have observed similar patterns in other livestock species,[65–69] suggesting common genomic features resulting from a close relationship with human. We observed significant selection signatures on loci related to reproduction. This is not unexpected given that chickens are one of the world's most important and efficient protein sources, and many chicken breeds have been developed with a significantly improved egg-laying capacity and a shorter time to maturity. At least for this trait, if not for all genes the timing of the origin of the selective sweeps likely significantly postdates the temporal origins of domestic populations.[57] In

addition, we found that TSHR-Gly558Arg, a previously proposed loci responsible for chicken domestication,[21] is nearly fixed in both chickens and *G. g. spadiceus*, but remains at low frequencies in other RJF subspecies (Supplementary information, Fig. S58). This result firstly suggests that the mutation arose prior to domestication, but only in *G. g. spadiceus*, and that this mutation must have been at high frequency in the original population of domestic chickens. The proportion of the 'wild type' allele likely increased as domestic birds moved west and admixed with subspecies of RJF that did not possess this missense mutation, before selection for the 'domestic' version of the gene drove it back to its modern ubiquity in domestic flocks. Alternatively, this allele could have been introgressed into *G. g. spadiceus* from domestic chickens and swept to high frequency. This hypothesis, however, seems less likely given that this allele is found at a very low frequency in other RJF subspecies that have also experienced gene flow from domestic chickens. Analyzing ancient genomes from chicken and RJF spanning a wide timeframe and range is expected to more precisely determine when the selection on these traits first began[18] as well as more precisely pinpoint the geographic and temporal origins and dispersal patterns of domestic chickens.

The novel findings from this study provide new insights into the origin and evolutionary history of domestic chickens. The identification of unique genomic landscapes of all RJF subspecies and three additional jungle fowl species suggests that conservation efforts should be made to safeguard them from extinction. These rich genomic resources will pave the way to facilitate ongoing explorations into the biocultural history of the relationship between humans and chickens as well as the development of fast-growing, high-quality and cost-effective lineages.

MATERIALS AND METHODS

Sample collection and genome sequencing
We collected 787 bird samples for whole-genome sequencing, including 627 domestic chickens, 142 RJFs (Supplementary information, Table S1; *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)), four Ceylon jungle fowls (*G. lafayettei*), two gray jungle fowls (*G. sonnerati*) and 12 green jungle fowls (*G. varius*). Total genome DNA was extracted and purified from blood or muscle of bird using phenol-chloroform method. Genome DNA for each sample was sheared into fragment of 300–600 bp using Covaris system (https://covaris.com/). Next-generation genome sequencing libraries were constructed according to standard protocol of library preparation kit. Genome sequencing was performed on the Illumina HiSeq and NextSeq platforms. 76 previously published genomes for chicken and RJFs[24,26–29] were integrated into our study.

Sequence alignment and variant calling
Raw sequencing reads were trimmed using Btrim software[70] to filter out low-quality bases and sequences. High-quality reads were aligned against the chicken reference genome (*Galgal4*) using bwa "BWA-MEM" algorithm.[71] Alignment bam/sam files were next subjected to a series of processing and filtrations including position sorting, duplicated read marking and removal, local realignment and base quality recalibration, which were carried out using tools available in Picard (http://picard.sourceforge.net) and Genome Analysis Toolkit (GATK) packages.[72] SNPs were genotyped and filtered using the UnifiedGenotyper and VariantFiltration tools in GATK package, respectively.

Phylogeny, PCA and structure analysis
Maximum-likelihood tree was built using FastTree program (version: 2.1.9; available at http://www.microbesonline.org/fasttree/)[73] based on whole-genome data. PCA was performed using both GCTA software[74] and smartPCA program from Eigensoft package (version: 5.0.2).[75] Genetic structure clustering

Article

was performed using ADMIXTURE program by assuming that the number of ancestral populations (K) increased gradually.[31] Ten independent runs with different random seed were analyzed, and these matrixes were summarized and compiled with CLUMPAK.[76] Genotypes for both PCA and admixture clustering were pruned based on linkage disequilibrium by PLINK.[77]

Population divergence and demographic estimations

The program Beagle[43] was used to impute the missing genotype and phase of genotypes into the haplotypes. Demographic history and population size fluctuation over time for RJF and chicken was inferred using MSMC.[34] Generation time (g) of one year and mutation rate (μ) of 1.91×10^{-9} substitutions per site per year were used to scale the MSMC estimations.[35] Splitting time for each population pairs was retrieved when the relative CCR drops to 50%. The 2-D unfolded site frequency spectrum (SFS) for each population pair was generated using a modified script from *dadi*.[78] All assumed demographic models were tested using fastsimcoal2 program.[40] For each model, 100 independent runs were performed with varying starting points.

Population admixture analysis

Outgroup f_3- and f_4-statistics were computed using the *threepop* and *fourpop* programs from TreeMix package,[32] respectively. Population splitting and admixture analyses were carried out using TreeMix program.[32] Admixture graphs were inferred using *qpGraph* program from AdmixTools package.[33,79] In these analyses, green jungle fowl was used as outgroup. Local ancestry inference was carried out using PCAdmix program (version: 1.0)[39] based on phased genotypes that were inferred by Beagle (version: 4.1).[43]

Genomic regions introgressed between green, gray and Ceylon jungle fowls and domestic chicken were inferred by Z-rIBD method[44] using sliding windows in the window size of 10 kb with 5 kb increment along each chromosome. Haplotype trees for these putatively introgressed fragments were constructed using MEGA7[80] to determine the direction of the introgression.

Scanning for selective sweeps

The LSBL statistics[45] and π-ratios[46] were used to identify signature of positive selection in domestic chickens. These analyses were performed using 50 kb sliding windows with a shifting increment of 25 kb at each step. Two sets of LSBL scores were computed: $LSBL_j$, by comparing chickens with *G. g. spadiceus* and *G. g. jabouillei*; and $LSBL_m$, by comparing chickens with *G. g. spadiceus* and *G. g. murghi*. π-ratio was computed based on $\pi_{G.g.spadiceus}/\pi_{chicken}$. We Z-transformed each statistic and applied a Z-score ≥ 3.3 (corresponding to $P \leq 0.001$) to retrieve putative selective sweeps. Genes encompassed within these genomic regions were annotated using Variant Effect Predictor (VEP).[81] Functional enrichment terms for these genes, including Gene Ontology (GO) categories, KEGG pathways, and Human Phenotype Ontologies (HPOs) were retrieved using g:Profiler.[82]

DATA AVAILABILITY

The raw sequencing data, alignment BAM files, and genotypes (VCF) have been deposited in the ChickenSD database (http://bigd.big.ac.cn/chickensd/).

ADDITIONAL INFORMATION

Supplementary information accompanies this paper at https://doi.org/10.1038/s41422-020-0349-y.

Competing interests: The authors declare no competing interests.

REFERENCES

1. Lawler, A. Why did the chicken cross the world? Atria Books, New York (2014).
2. Larson, G. & Fuller, D. Q. The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).
3. Peters, J. et al. Questioning new answers regarding Holocene chicken domestication in China. *Proc. Natl Acad. Sci. USA* **112**, E2415 (2015).
4. Tixier-Boichard, M., Bed'hom, B. & Rognon, X. Chicken domestication: from archeology to genomics. *C. R. Biol.* **334**, 197–204 (2011).
5. Miao, Y. W. et al. Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity* **110**, 277–282 (2013).
6. Liu, Y. P. et al. Multiple maternal origins of chickens: out of the Asian jungles. *Mol. Phylogenet. Evol.* **38**, 12–19 (2006).
7. Fumihito, A. et al. Monophyletic origin and unique dispersal patterns of domestic fowls. *Proc. Natl Acad. Sci. USA* **93**, 6792–6795 (1996).
8. Lawler, A. Animal domestication. In search of the wild chicken. *Science* **338**, 1020–1024 (2012).
9. Peters, J., Lebrasseur, O., Deng, H. & Larson, G. Holocene cultural history of red jungle fowl (Gallus gallus) and its domestic descendant in East Asia. *Quat. Sci. Rev.* **142**, 102–119 (2016).
10. Eda, M. et al. Reevaluation of early Holocene chicken domestication in northern China. *J. Archaeol. Sci.* **67**, 25–31 (2016).
11. Girdland Flink, L. et al. Establishing the validity of domestication genes using DNA from ancient chickens. *Proc. Natl Acad. Sci. USA* **111**, 6184–6189 (2014).
12. Wang, G. D., Xie, H. B., Peng, M. S., Irwin, D. & Zhang, Y. P. Domestication genomics: evidence from animals. *Annu. Rev. Anim. Biosci.* **2**, 65–84 (2014).
13. Larson, G. & Burger, J. A population genetics view of animal domestication. *Trends Genet.* **29**, 197–205 (2013).
14. Wang, G. D. et al. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* **26**, 21–33 (2016).
15. Ní Leathlobhair, M. et al. The evolutionary history of dogs in the Americas. *Science* **361**, 81–85 (2018).
16. Frantz, L. A. et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**, 1228–1231 (2016).

17. Chen, N. et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 2337 (2018).

18. Fages, A. et al. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419–1435 (2019).

19. Flowers, J. M. et al. Cross-species hybridization and the origin of North African date palms. *Proc. Natl Acad. Sci. USA* **116**, 1651–1658 (2019).

20. Li, D. et al. Population genomics identifies patterns of genetic diversity and selection in chicken. *BMC Genom.* **20**, 263 (2019).

21. Rubin, C. J. et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591 (2010).

22. Wang, M. S. et al. An evolutionary genomic perspective on the breeding of dwarf chickens. *Mol. Biol. Evol.* **34**, 3081–3088 (2017).

23. Guo, Y. et al. A complex structural variation on chromosome 27 leads to the ectopic expression of HOXB8 and the Muffs and beard phenotype in chickens. *PLoS Genet.* **12**, e1006071 (2016).

24. Wang, M. S. et al. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. *Cell Res.* **26**, 556–573 (2016).

25. Guo, X. et al. Whole-genome resequencing of Xishuangbanna fighting chicken to identify signatures of selection. *Genet. Sel. Evol.* **48**, 62 (2016).

26. Wang, M. S. et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol. Biol. Evol.* **32**, 1880–1889 (2015).

27. 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).

28. Yi, G. et al. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genom.* **15**, 962 (2014).

29. Fan, W. L. et al. Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol. Evol.* **5**, 1376–1392 (2013).

30. Huang, X. H. et al. Was chicken domesticated in northern China? New evidence from mitochondrial genomes. *Sci. Bull.* **63**, 743–746 (2018).

31. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

32. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).

33. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

34. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).

35. Nam, K. et al. Molecular evolution of genes in avian genomes. *Genome Biol.* **11**, R68 (2010).

36. Collias, N. E. & Saichuae, P. Ecology of the red jungle fowl in Thailand and Malaya with reference to the origin of domestication. *Nat. Hist. Bull. Siam. Soc.* **22**, 189–209 (1967).

37. Xiang, H. et al. Early Holocene chicken domestication in northern China. *Proc. Natl Acad. Sci. USA* **111**, 17564–17569 (2014).

38. West, B. & Zhou, B. X. Did chickens go North? New evidence for domestication. *J. Archaeol. Sci.* **15**, 515–533 (1988).

39. 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**, 343–364 (2012).

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. Eriksson, J. et al. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genet.* **4**, e1000010 (2008).

42. Lawal, R. A. et al. The wild species genome ancestry of domestic chickens. *BMC Biol.* **18**, 13 (2020).

43. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).

44. Bosse, M. et al. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* **5**, 4392 (2014).

45. Shriver, M. D. et al. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum. Genom.* **1**, 274–286 (2004).

46. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273 (1979).

47. Librado, P. et al. Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442–445 (2017).

48. Rohner, N. et al. Duplication of fgfr1 permits Fgf signaling to serve as a target for selection during domestication. *Curr. Biol.* **19**, 1642–1647 (2009).

49. Atikuzzaman, M. et al. Selection for higher fertility reflects in the seminal fluid proteome of modern domestic chicken. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **21**, 27–40 (2017).

50. Schutz, K. E. et al. Major growth QTLs in fowl are related to fearful behavior: possible genetic links between fear responses and production traits in a red junglefowl x White Leghorn intercross. *Behav. Genet.* **34**, 121–130 (2004).

51. Schutz, K. et al. QTL analysis of a red junglefowl x White Leghorn intercross reveals trade-off in resource allocation between behavior and production traits. *Behav. Genet.* **32**, 423–433 (2002).

52. Bédécarrats, Y. G., Shimizu, M. & Guémené, D. Gonadotropin releasing hormones and their receptors in avian species. *J. Poult. Sci.* **43**, 199–214 (2006).

53. Sharp, P. J. et al. Physiological roles of chicken LHRH-I and -II in the control of gonadotrophin release in the domestic chicken. *J. Endocrinol.* **124**, 291–299 (1990).

54. Marques, P. et al. Physiology of GNRH and gonadotropin secretion. MDText.com, Inc. South Dartmouth, Massachusetts (2000).

55. Gandhi, R. et al. The *Drosophila* kinesin-like protein KLP67A is essential for mitotic and male meiotic spindle assembly. *Mol. Biol. Cell* **15**, 121–131 (2004).

56. Liu, X. S. et al. Germinal cell aplasia in Kif18a mutant male mice due to impaired chromosome congression and dysregulated BubR1 and CENP-E. *Genes Cancer* **1**, 26–39 (2010).

57. Loog, L. et al. Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Mol. Biol. Evol.* **34**, 1981–1990 (2017).

58. Qanbari, S. et al. Genetics of adaptation in modern chicken. *PLoS Genet.* **15**, e1007989 (2019).

59. Ye, W. et al. Early middle Holocene climate oscillations recorded in the Beihuqiao Core, Yuhang, Zhejiang Province, China. *J. Paleolimnol.* **59**, 263–278 (2017).

60. Wu, M. S. et al. Holocene hydrological and sea surface temperature changes in the northern coast of the South China Sea. *J. Asian Earth Sci.* **135**, 268–280 (2017).

61. Fan, Z. et al. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.* **26**, 163–173 (2016).

62. Wurster, C. M. et al. Forest contraction in north equatorial Southeast Asia during the Last Glacial Period. *Proc. Natl Acad. Sci. USA* **107**, 15508–15511 (2010).

63. Mellars, P. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* **313**, 796–800 (2006).

64. Fumihito, A. et al. One subspecies of the red junglefowl (Gallus gallus gallus) suffices as the matriarchic ancestor of all domestic breeds. *Proc. Natl Acad. Sci. USA* **91**, 12505–12509 (1994).

65. Alberto, F. J. et al. Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* **9**, 813 (2018).

66. 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).

67. Wang, G. D. et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.* **4**, 1860 (2013).

68. Daly, K. G. et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**, 85–88 (2018).

69. Schubert, M. et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl Acad. Sci. USA* **111**, E5661–E5669 (2014).

70. Kong, Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **98**, 152–153 (2011).

71. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843–2851 (2014).

72. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

73. 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**, 1641–1650 (2009).

74. 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).

75. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

76. 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**, 1179–1191 (2015).

77. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

78. 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).

79. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).

80. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

81. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

82. Reimand, J., Arak, T. & Vilo, J. g:Profiler-a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.* **39**, W307–W315 (2011).