# UC Santa Cruz
## UC Santa Cruz Previously Published Works

**Title**
Large-scale genomic analysis reveals the genetic cost of chicken domestication

**Permalink**
https://escholarship.org/uc/item/7ft262q7

**Journal**
BMC Biology, 19(1)

**ISSN**
1478-5854

**Authors**
Wang, Ming-Shan
Zhang, Jin-Jin
Guo, Xing
et al.

**Publication Date**
2021-12-01

**DOI**
10.1186/s12915-021-01052-x

Peer reviewed

**BMC Biology**

**RESEARCH ARTICLE**                                                                                    **Open Access**

# Large-scale genomic analysis reveals the genetic cost of chicken domestication

Ming-Shan Wang[1,2,3,4†] (iD), Jin-Jin Zhang[1,2†], Xing Guo[5†], Ming Li[6†], Rachel Meyer[4†], Hidayat Ashari[7,8†],
Zhu-Qing Zheng[9], Sheng Wang[1,2], Min-Sheng Peng[1,2], Yu Jiang[6], Mukesh Thakur[1,10],
Chatmongkon Suwannapoom[11,12], Ali Esmailizadeh[1,13], Nalini Yasoda Hirimuthugoda[1,14],
Moch Syamsul Arifin Zein[7], Szilvia Kusza[15], Hamed Kharrati-Koopaee[13,16], Lin Zeng[1,2], Yun-Mei Wang[17],
Ting-Ting Yin[1,2], Min-Min Yang[1,2], Ming-Li Li[1,2], Xue-Mei Lu[1,2,18], Emiliano Lasagna[19], Simone Ceccobelli[19],
Humpita Gamaralalage Thilini Nisanka Gunwardana[14], Thilina Madusanka Senasig[14], Shao-Hong Feng[1,20],
Hao Zhang[21], Abul Kashem Fazlul Haque Bhuiyan[22], Muhammad Sajjad Khan[23],
Gamamada Liyanage Lalanie Pradeepa Silva[24], Le Thi Thuy[25], Okeyo A. Mwai[26],
Mohamed Nawaz Mohamed Ibrahim[26], Guojie Zhang[1,18,27,28], Kai-Xing Qu[29], Olivier Hanotte[30,31], Beth Shapiro[3,4],
Mirte Bosse[32*], Dong-Dong Wu[1,2,18*], Jian-Lin Han[8,26*] and Ya-Ping Zhang[1,2,18,33*]

## Abstract

**Background:** Species domestication is generally characterized by the exploitation of high-impact mutations through processes that involve complex shifting demographics of domesticated species. These include not only inbreeding and artificial selection that may lead to the emergence of evolutionary bottlenecks, but also post-divergence gene flow and introgression. Although domestication potentially affects the occurrence of both desired and undesired mutations, the way wild relatives of domesticated species evolve and how expensive the genetic cost underlying domestication is remain poorly understood. Here, we investigated the demographic history and genetic load of chicken domestication.

* Correspondence: mirte.bosse@wur.nl; wudongdong@mail.kiz.ac.cn;
h.jianlin@cgiar.org; zhangyp@mail.kiz.ac.cn
†Ming-Shan Wang, Jin-Jin Zhang, Xing Guo, Ming Li, Rachel Meyer and
Hidayat Ashari contributed equally to this work.
32Wageningen University & Research - Animal Breeding and Genomics, 6708
PB Wageningen, The Netherlands
1State Key Laboratory of Genetic Resources and Evolution, Yunnan
Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of
Zoology, Chinese Academy of Sciences, Kunming 650223, China
8CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources,
Institute of Animal Science, Chinese Academy of Agricultural Sciences
(CAAS), Beijing 100193, China
Full list of author information is available at the end of the article

Wang *et al. BMC Biology*      (2021) 19:118

Page 2 of 16

**Results:** We analyzed a dataset comprising over 800 whole genomes from both indigenous chickens and wild jungle fowls. We show that despite having a higher genetic diversity than their wild counterparts (average $\pi$, 0.00326 vs. 0.00316), the red jungle fowls, the present-day domestic chickens experienced a dramatic population size decline during their early domestication. Our analyses suggest that the concomitant bottleneck induced 2.95% more deleterious mutations across chicken genomes compared with red jungle fowls, supporting the "cost of domestication" hypothesis. Particularly, we find that 62.4% of deleterious SNPs in domestic chickens are maintained in heterozygous states and masked as recessive alleles, challenging the power of modern breeding programs to effectively eliminate these genetic loads. Finally, we suggest that positive selection decreases the incidence but increases the frequency of deleterious SNPs in domestic chicken genomes.

**Conclusion:** This study reveals a new landscape of demographic history and genomic changes associated with chicken domestication and provides insight into the evolutionary genomic profiles of domesticated animals managed under modern human selection.

**Keywords:** Domestication, Bottleneck, Genetic load, Deleterious mutation, Domestic chicken

## Background

All organisms carry a certain level of deleterious mutations in their genomes, which can potentially affect their fitness [1, 2]. The majority of these harmful mutations are detrimental and recessive—only a few are dominant or recessive lethal [3]. Such mutations can best be seen as high-impact mutations (that is, affecting the functioning or expression of a gene) [4, 5]. Most of them have a negative effect; however, some may result in a desirable phenotype and are therefore maintained by natural and artificial selection. Some of these alleles that are preferred in an artificial breeding setting would nevertheless be detrimental in the wild. The evolution of domestic species is characterized by the exploitation of high-impact mutations during inbreeding, artificial selection, and post-divergence gene flow [6–10], which could affect the occurrence of both desired and undesired high-impact mutations [11]. Extensive studies have reported that domestic species, such as horses [12], dogs [13], rice [14, 15], sheep [10], and tomatoes [16], are burdened by many more deleterious mutations than their wild relatives. The "cost of domestication" hypothesis was proposed to explain this general pattern observed in these domestic species [17]. It suggests that bottlenecks along with domestication reduced the power of purifying selection to remove deleterious variants, therefore resulting in a dramatic accumulation of deleterious variants in domesticated species. However, this model cannot be generalized to all major domesticates because some of them lack a domestication bottleneck [18, 19]. For example, genomic assessment of pigs [20], bees [21], and some crops [18] exhibited no significant historical decline in their genetic diversity relative to respective wild progenitors.

Chicken is believed to have been domesticated from the red jungle fowl (RJF). Early genomic studies have identified a number of variants that differentiate chicken from RJF, facilitating our understanding of the genetic changes underlying chicken domestication [22, 23]. It was thought that chicken has been domesticated via a commensal pathway within the Holocene, in which its early domestication was assumed to be fully unintentional [8] and not marked by a bottleneck [19]. However, some studies proposed a contradictory opinion based on the observations of evaluated high-impact mutations in specific chicken breeds [2, 24], although the establishment of these fancy and commercial birds has likely resulted in a more recent bottleneck. To date, whether there was a domestication bottleneck, its strength and effect, and how it potentially shaped the pattern of high-impact mutations in relation to other factors, including admixture and selection, are unresolved in domestic chicken.

Measuring the magnitude of domestication bottleneck by direct comparison of genetic diversity in current domesticates with their true wild progenitors is often impossible (e.g., extinction of the ancestors of cattle and horses) or complicated by several factors [8, 25, 26]. First, accurate identification of the sole ancestor(s) of domesticated species remains very challenging because the potential wild relatives either continue to evolve as closely related genetic entities with a large geographic range (e.g., RJFs and gray wolves) [27–29] or survive only in very small fragmented populations (e.g., sheep, goats, and buffaloes) [30, 31] due to overhunting among other human activities. Second, hybridization between domesticated and wild populations is common in nature and thus could mislead the estimation of their genetic differences [32, 33]. Domesticated chickens have been affected by gene flow via hybridization with other RJFs and jungle fowl species over thousands of years [34, 35]. RJFs are widely distributed and could be assigned into five subspecies (*G. g. spadiceus*, *G. g. murghi*, *G. g. jabouillei*, *G. g. gallus*, and *G. g. bankiva*) ranging across

South and Southeast Asia where they may have been hybridizing with village hens [36]. The third is the afore-mentioned issue that wild species can rarely escape from climate and anthropogenic pressures, and this creates somewhat parallel trajectories to evolution under domestication [28, 37].

In the first phase of our 1K Chicken Genomes Project (1K CGP; https://bigd.big.ac.cn/chickensd/), through sequencing and analyzing 863 genomes from both jungle fowls and indigenous chickens sampled across South, Southeast, and East Asia as well as Europe (including 149 RJFs covering all five subspecies sampled in their natural ranges), we demonstrated that all domestic chickens were monophyletic, derived from an RJF lineage of *G. g. spadiceus* (GGS) whose present-day range is predominantly in northern Thailand, southwestern China, and Myanmar, and then regained genetic diversity via introgression from additional RJF subspecies and jungle fowl species during

their dispersals out of the domestication center [27]. In this study, based on the new knowledge and genomes (including 696 domestic chickens spanning Eurasia and 45 GGS from Thailand and Yunnan Province, China), we examined the demographic history for chickens before and after their domestication, investigated the distribution and frequency of high-impact mutations across their genomes, and estimated the genetic cost of chicken domestication and breed formation.

## Results

### Evidence of bottleneck in chicken domestication

To have baseline information on the genomic diversity for present-day domestic chickens and GGS, we estimated nucleotide diversity (π) for each population. The average π for all domestic chickens was 3.26e−3, slightly higher than that for GGS (mean 3.16e−3; $P < 2.2e-16$, Wilcoxon signed-rank test; Fig. 1a). We note that this
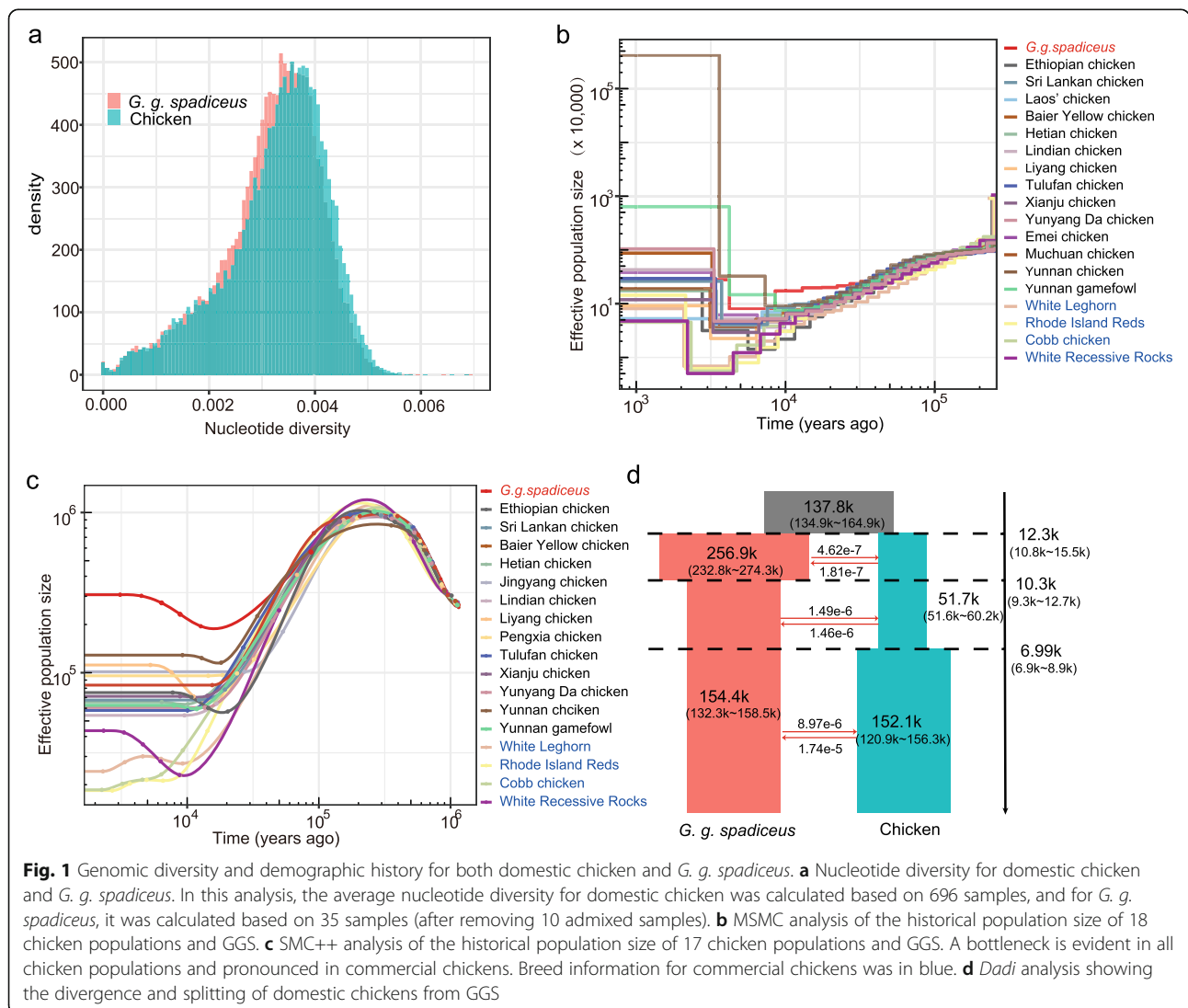


**Fig. 1** Genomic diversity and demographic history for both domestic chicken and *G. g. spadiceus*. **a** Nucleotide diversity for domestic chicken and *G. g. spadiceus*. In this analysis, the average nucleotide diversity for domestic chicken was calculated based on 696 samples, and for *G. g. spadiceus*, it was calculated based on 35 samples (after removing 10 admixed samples). **b** MSMC analysis of the historical population size of 18 chicken populations and GGS. **c** SMC++ analysis of the historical population size of 17 chicken populations and GGS. A bottleneck is evident in all chicken populations and pronounced in commercial chickens. Breed information for commercial chickens was in blue. **d** *Dadi* analysis showing the divergence and splitting of domestic chickens from GGS

Wang *et al. BMC Biology* (2021) 19:118

Page 4 of 16

result should be interpreted with caution because our sampling of GGS likely did not cover all their genetic diversity while our chicken samples were from a wide range sampling spanning Eurasia. Pooling analysis of genomes from these diverse chicken populations together possibly inflates the estimation of genetic diversity (Additional file 1: Figure S1).

We used the pair-wise sequential Markovian coalescent (PSMC) [38] approach to estimate the historical effective population sizes (Ne) of domestic chickens and GGS. As the efficiency of this analysis relies on heterozygotes across genomes [39], the result from low-coverage genomes is unreliable. The number of high-coverage genomes in 1K CGP is limited [27], so we also included recently published chicken genomes [35, 40, 41] with a coverage of ≥ 20-folds in this analysis (Additional file 1: Table S1). Our data contain genomes from GGS and 18 chicken populations including commercial breeds (White Leghorn, White Recessive Rocks, Rhode Island Red, and Cobb chicken), Ethiopian, Sri Lankan, Laos, and Chinese local chickens. The estimations were scaled by a mutation rate of 1.91e−9 substitution per site per year and a generation time of 1 year [42]. PSMC revealed that both domestic chickens and GGS had nearly identical demographic histories before 20 thousand years ago (kya) (Additional file 1: Figure S2), which is expected as the chicken was originated from GGS [27]. Specifically, initialed at 1 million years ago (Mya), Ne for chicken and GGS showed expansion and reached a maximum of around 100 kya, followed by continuous contraction until 20 kya. The cycles of past population expansions and contractions were similar to those observed in other wild birds, supporting the claim that climate fluctuations during the Quaternary have likely shaped the evolution and speciation of many bird species [43].

However, PSMC has limited power to reveal the recent Ne within 10 kya. We then used MSMC [44], a method similar to PSMC, which could analyze genomes from more than one individual for each population and therefore has an enhanced power to infer the demographic changes for relatively recent evolutionary events like domestication. Our estimations were based on four haplotype genomes (two individuals) for each population. Consistent with PSMC, MSMC revealed a comparable and dramatic Ne contraction for 18 chicken populations and GGS between 100 and 20 kya (Fig. 1b). Thereafter, chicken and GGS showed obvious differentiation. Specifically, Ne for GGS remained relatively constant before the rapid decline onset ~8 kya but then stabilized ~4 kya, while domestic chickens showed a continuous Ne decline until Yunnan local chicken and Yunnan game fowl recovered ~6–7 kya and other chicken populations recovered later ~2–4 kya. Commercial chicken breeds, including White Recessive Rocks, Cobb chicken, White Leghorn,

and Rhode Island Red, have much smaller Ne than other chickens, and their Ne recovered much later (~2 kya), consistent with the fact that commercial chickens have been subject to intensified artificial selection and inbreeding. It should be noted that more recent Ne estimated by MSMC tends to be inflated [44], especially in analyzing more than two haplotypes. Here, we further inferred historical Ne for GGS and each chicken population using SMC++ [45], a method estimating demographic histories based on multiple genomes without phasing. For each population, we allowed five genomes with sequencing coverage over 15-folds, yielding a total of 17 chicken populations for the analysis from our current dataset (Fig. 1c). This analysis also revealed strong evidence of domestication bottleneck for chicken compared with GGS, broadly consistent with the result from MSMC. Compared with MSMC, SMC++ has a higher resolution for estimating recent population histories. For example, SMC++ analyses also revealed that commercial chickens (White Leghorn, White Recessive Rock, Rhode Island Red, and Cobb chicken) have a stronger bottleneck and much smaller recent Ne compared with other chickens. Ne for Yunnan local chicken, Liyang chicken, White Recessive Rocks, Sri Lankan local chicken, Ethiopian local chicken, and game fowls showed a recent recovery after the bottleneck during 10 kya.

Because PSMC, MSMC, and SMC++ do not take into account the possible effects of admixture [46, 47], the estimated initiation of Ne differentiation does not necessarily correlate with the splitting time. To further infer the evolutionary history of domestic chickens, we used *dadi* [48] to fit the joint site frequency spectrum (SFS) between domestic chicken and GGS populations. We tested four assumed demographic models (Additional file 1: Figure S3-S4 and Tables S2-S3) and found that model 3 had the highest likelihood (Fig. 1d), suggesting its best fit to observed SFS. Under this model, we estimated that modern domestic chickens and GGS separated from each other ~12,300 (95% confident intervals (CI)10,800–15,500) years ago. After that, GGS showed a slight expansion in Ne until 10,300 years ago (95% CI 9300–12,700), followed by a contraction from around 257,000 down to 154,000 birds, which was likely resulted from recent habitat loss and overhunting. On the contrary, the average Ne for domestic chickens dramatically declined from around 138,000 birds on their separation from GGS down to 52,000 until ~6990 (6900–8900) years ago when they started to be recovered up to ~152,000 birds till today. MSMC, SMC++, and *dadi* analyses suggested that domestic chickens experienced a continuous Ne decline since their splitting off from GGS, indicating a bottleneck in chicken domestication. We noted that admixture with local jungle fowl or other domestic chicken populations is pervasive [27], which likely affects

Wang *et al. BMC Biology*     (2021) 19:118

Page 5 of 16

the inference of demographic history [49], so the bottleneck time we estimated here is inconclusive.

### Identification of high-impact mutations

To measure whether the domestication-associated bottleneck could have induced the rise of high-impact SNPs (hSNPs) in chicken genomes, we analyzed the variants called in domestic chicken and GGS genomes among our 1K CGP. We found that protein-coding regions accounted for ~4.2% of the chicken genome, and 1.6% of the genomic variants (435,919) were present in these exonic regions (Fig. 2a), which were nearly two times higher than those observed in other domestic animals including dogs, pigs, cattle, and horses (Additional file 1: Table S4). These exonic variants were only differentiated slightly between domestic chickens and GGS, suggesting that they have been subjected to evolutionary constraints. We classified mutations from the exonic regions into non-synonymous and synonymous substitutions and identified 146,193 non-synonymous SNPs, accounting for around 33.5% of the total exonic SNPs. The estimated ratio for the numbers of non-synonymous SNPs over exonic SNPs was similar to those observed in other domestic animals (Additional file 1: Table S4). To assess their potential effects (i.e., tolerant or deleterious) on associated amino acid changes in protein sequences of domestic chickens, we performed a PROVEAN [50] analysis. Based on the score threshold ≤ −2.5, we detected 22,282 potential hSNPs (Fig. 2b). Compared to mammals, bird chromosomes are highly variable in size. Chicken chromosomes are classified into three classes including 5 macrochromosomes (chrs: 1–5), 28 microchromosomes (chrs: 11–38), and 4 intermediate chromosomes (chrs: 6–10) [51]. We compared the PROVEAN scores of non-synonymous mutations and found that the average PROVEAN scores (damaging effect) of variants on the microchromosomes were significantly lower than those on the remaining chromosomes (Additional file 1: Figure S5). This finding is consistent with the previous report that microchromosomes have been subjected to evolutionary constraint or more efficient purifying selection because of their higher recombination rates [51], and therefore, mutations in conserved regions are more likely to be harmful.

To evaluate the potential biological roles of the associated genes, we retrieved genes carrying the non-synonymous SNPs of PROVEAN score < −10 and performed functional enrichment analyses including GO and Human Phenotype Ontology (HPO). A total of 346 protein-coding genes were retrieved, and they are involved in multiple functional GO and HPO categories, including abnormalities of nervous system physiology, growth, bone and muscle development and morphology, cardiovascular and respiratory system, metabolism/homeostasis, vision,

and immunity (Additional file 1: Table S5, Fig. 2c). These mutations were at low frequencies in domestic chickens, most of them are likely linked to health problems faced by modern poultry industries, as highly productive birds have been suffering from brittle bones, blindness, crippling leg disorders, ascites (a disease of the lungs and heart), and sudden death syndrome [52, 53]. Nevertheless, a few of them may be the target of positive selection during domestication and/or recent genetic improvement or breeding for specific traits.

### Validation of the function of *TSHR*-Gly558Arg using a transgenic mouse model

In an early investigation [22], thyroid-stimulating hormone receptor (*TSHR*) showed the strongest signal of selection, with one missense mutation (chr5:40,089, 599G/A; *TSHR*-Gly558Arg) being nearly fixed in domestic chickens compared to RJFs (< 23%; unclear subspecies classification). A recent paleogenetic study, however, showed the timing of selection on this gene in ancient and modern European chickens and concluded that the dramatic rise in frequency to modern ubiquity only began 1.1 kya [54]. Our recent study showed that this mutation was nearly fixed (> 90%) in both indigenous chickens and GGS but maintained at extremely low frequencies among other RJF subspecies [27]. Should any claims be based on the knowledge that this mutation is functional, however, the general biological role of this gene and the specific functional consequence of this mutation have not been resolved in domestic chickens. In our analysis, the *TSHR*-Gly558Arg has a PROVEAN score at −6.981, to be indicative of potentially high effect (< −2.5). Therefore, it is interesting to assess the potential biological function of this mutation as a potential proof of concept to test all our predictions.

Because of the challenge in editing an avian genome efficiently and precisely, several studies have employed transgenic mouse or human cell lines [55] or zebrafish [56, 57] models to test the potential roles of specific mutations of interest identified in the bird genomes. As *TSHR* is associated with development and metabolism in mice [58] and glycine at this position in chicken is conserved among all known vertebrate TSHR amino acid sequences [22, 59], we constructed chicken *TSHR*-558Arg knock-in mice (matched with mice *TSHR*-559Arg) to test whether this mutation has any biological effects (Additional file 1: Figure S6 and Table S6). At normal conditions, *TSHR*-559Arg homozygous mice developed significantly smaller ($P < 0.01$) bodies in both sexes than the wild-type mice (Fig. 3a, b, Additional file 1: Figure S7). At 30 °C, 18 °C, and 5 °C experimental conditions, we measured the physical movements of the mice by counting their locomotor activities and found no significant difference ($P > 0.05$) between the *TSHR*-558Arg
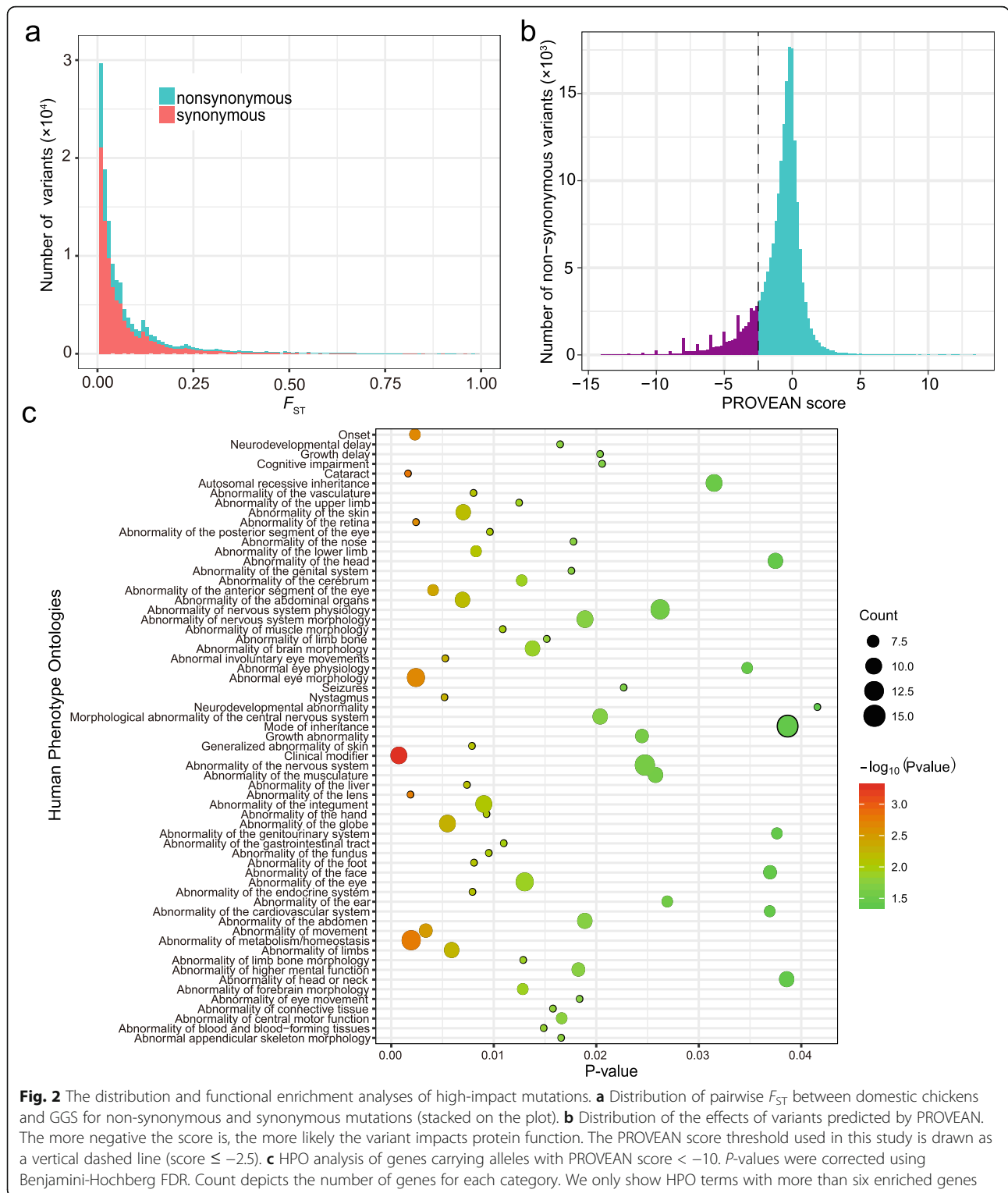
Wang *et al. BMC Biology* (2021) 19:118

Page 6 of 16



**Fig. 2** The distribution and functional enrichment analyses of high-impact mutations. **a** Distribution of pairwise $F_{ST}$ between domestic chickens and GGS for non-synonymous and synonymous mutations (stacked on the plot). **b** Distribution of the effects of variants predicted by PROVEAN. The more negative the score is, the more likely the variant impacts protein function. The PROVEAN score threshold used in this study is drawn as a vertical dashed line (score ≤ −2.5). **c** HPO analysis of genes carrying alleles with PROVEAN score < −10. *P*-values were corrected using Benjamini-Hochberg FDR. Count depicts the number of genes for each category. We only show HPO terms with more than six enriched genes

homozygous and wild-type mice (Fig. 3c). We also recorded their energy expenditures and metabolism rates and realized that the *TSHR*-559Arg homozygous mice had less food uptake (Additional file 1: Figure S8) and significantly lower oxygen ($VO_2$), calorie consumptions, and carbon dioxide production ($VCO_2$) than the wild-type mice (Fig. 3d–f; $P < 0.05$). This result suggests that *TSHR*-Gly558Arg is biologically functional in chickens. However, such validation for a single mutation cannot stand the effect of all mutations; further experimental
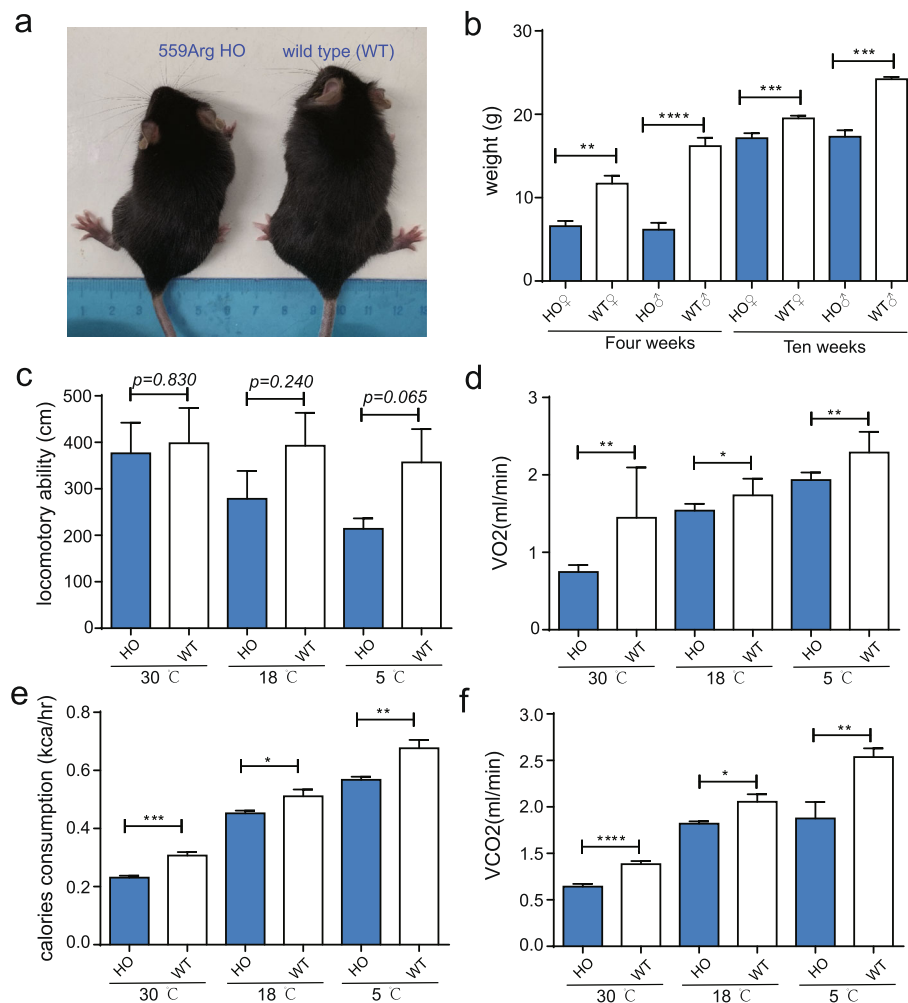
Wang *et al. BMC Biology*    (2021) 19:118

Page 7 of 16



**Fig. 3** Testing the function of *TSHR*-Gly559Arg using transgenic mouse model assay. **a** Photograph showing *TSHR*-559Arg knock-in homozygous (HO) and wild-type (WT) mice at 4 months old. **b** Bar plot shows that HO mice have significantly lower body weight than wild-type mice. **c** No difference in total locomotive ability between HO and wild-type mice. **d–f** HO mice have significantly lower oxygen consumption ($VO_2$), calorie consumption, and carbon dioxide exhalation ($VCO_2$) compared to wild-type mice. In **b**, $n = 6$ HO and $n = 8$ WT female and $n = 8$ HO and $n = 8$ WT male 4-week-old mice, as well as $n = 7$ HO and $n = 7$ WT female and $n = 12$ HO and $n = 14$ WT male 10-week-old mice were analyzed. In **c–f**, $n = 8$ for both HO and WT male mice were analyzed for each test. *$P < 0.05$; **$P < 0.01$;***$P < 0.001$; ****$P < 0.0001$. Statistical significance was measured by Student's t-test (two-tailed)

validation for additional and specific variants is warranted.

### Increased genetic load in domestic chickens

To compare the levels of genetic loads between domestic chickens and GGS, we evaluated the numbers and frequencies of non-synonymous and synonymous SNPs as well as the hSNPs among their genomes (Fig. 4). Our result showed that each chicken carried approximately 2.95% more hSNPs than GGS across their genomes ($P = 0.01865$, Wilcoxon signed-rank test; Fig. 4). Domestic chickens also had a significantly higher ratio of hSNPs relative to synonymous SNPs ($P = 1.83e−6$, Wilcoxon signed-rank test; Fig. 4). In particular, the average allele

frequency of hSNPs was significantly higher in domestic chickens than in GGS ($P < 2.2e−16$, Wilcoxon signed-rank test; Fig. 4). Our previous study suggests that, after originating from GGS, domestic chickens were further admixed with other jungle fowls during their dispersal out the domestication center [27]. The magnitude of gene flow is greatest between chickens and local jungle fowls, which might bias the estimation of genetic load. We further measured the number and ratio of hSNPs in some potentially drifted and/or isolated populations with no jungle fowl distributed in their natural ranges. To maximize more samples for each population, we chose Tibetan chicken, Beijing You chicken, Silkie chicken, chickens from Xinjiang province of China, and the
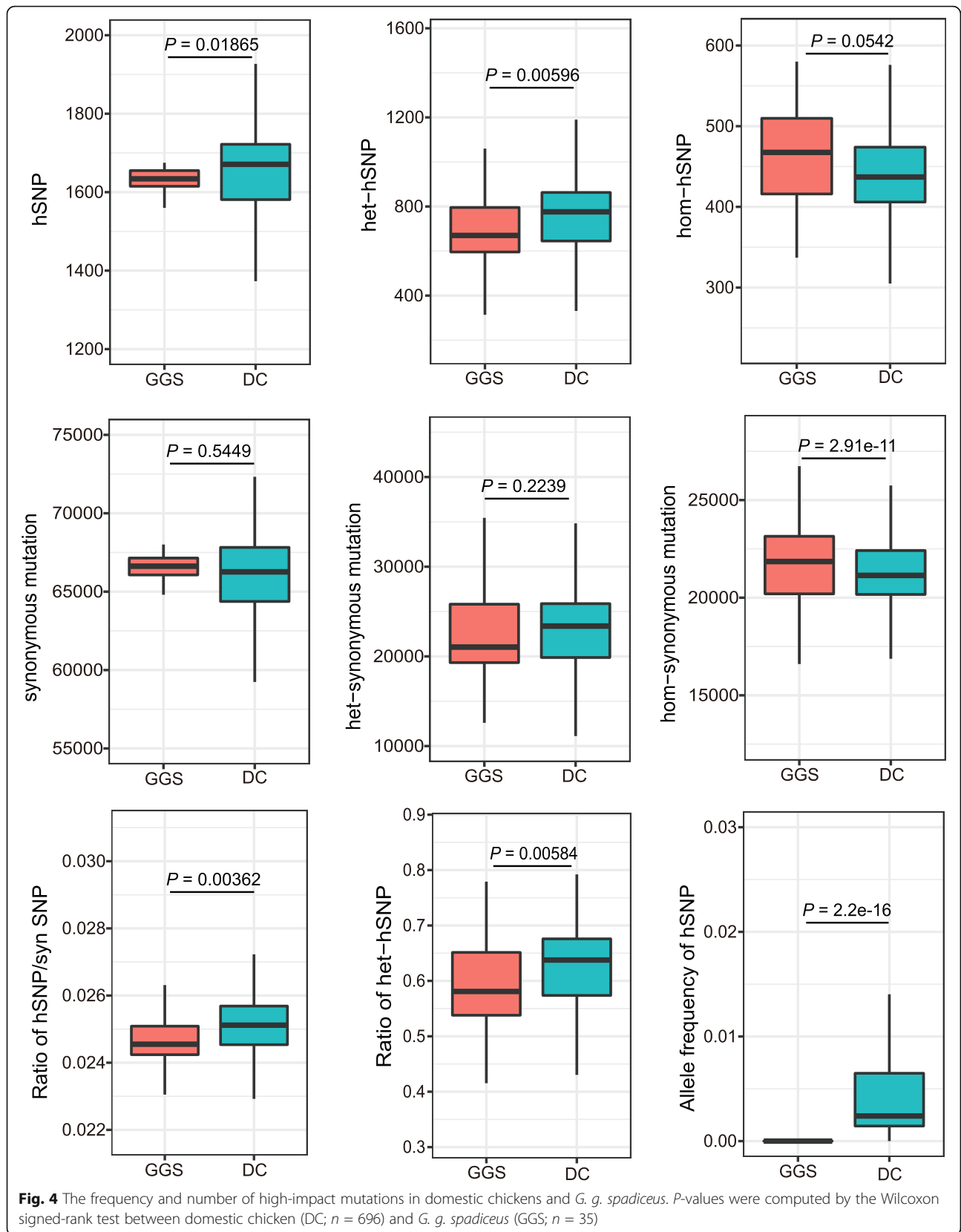
**Fig. 4** The frequency and number of high-impact mutations in domestic chickens and *G. g. spadiceus*. P-values were computed by the Wilcoxon signed-rank test between domestic chicken (DC; *n* = 696) and *G. g. spadiceus* (GGS; *n* = 35)

White Leghorn chicken for comparison. The ratio of deleterious mutation relative to synonymous mutation and the level of heterozygous deleterious mutation varied among these chicken populations but were all higher than GGS (Additional file 1: Figure S9). Collectively, these analyses suggest that domestication has led to a rapid accumulation of high-impact mutations, and thus, the genetic burden/load defined by the hSNPs was likely increased in domestic chickens.

We next compared the levels of hSNPs in both homozygous and heterozygous states and observed 62.4% hSNPs in domestic chickens to be maintained in heterozygote states, significantly higher than that in GGS (57.8%; $P$ = 0.00584, Wilcoxon signed-rank test; Fig. 4). In addition, domestic chickens carried far more heterozygous hSNPs ($P$ = 0.00596, Wilcoxon signed-rank test) but less homozygous hSNPs ($P$ = 0.0542, Wilcoxon signed-rank test; Fig. 4) than GGS (Fig. 4). However, the number of synonymous alleles, and these in the heterozygous states per genome, was comparable between domestic chickens and GGS ($P$ = 0.5449 and 0.2239, Wilcoxon signed-rank tests). The total number of homozygous synonymous alleles was higher in GGS than in domestic chickens ($P$ = 2.91e−11, Wilcoxon signed-rank test; Fig. 4). Similarly, we measured the level of heterozygous hSNPs in Tibetan chicken, Beijing You chicken, Silkie chicken, chickens from Xinjiang province of China, and the White Leghorn chicken and found these populations also had a higher ratio and number of heterozygous hSNPs than GGS (Additional file 1: Figure S9). These results suggest that the heterozygous mutation load was elevated in domestic chickens.

### The pattern of high-impact mutations in selective sweeps

How does selection affect the occurrence of hSNPs in domestic chicken genomes? To explore this, we retrieved the putatively selective sweeps identified by locus-specific branch length (LSBL) and π-ratio from our previous study [27] and compared the distribution pattern of hSNPs mapped within these sweeps with that in the remaining chicken genomic regions. Each of three sets of selective sweeps defined by LSBL(chicken, GGS, *G. g. jabouillei*) or LSBL(chicken, GGS, *G. g. murghi*) or π-ratio analysis possessed lower numbers of hSNPs (Additional file 1: Figure S10). This finding mirrors the observations in cassava [60] and grape [61], suggesting that the genes under selection tend to delimit hSNPs and/or to favor haplotypes carrying fewer hSNPs following the domestication and dispersal processes [60–62]. In addition, all the selective sweeps identified by each of the three statistics showed higher frequencies of hSNPs in domestic chickens than in GGS (Additional file 1: Figure S10). Some of these hSNPs may be the targets of selection to confer advantages in phenotypic or adaptive

innovations, while most of them were likely a result of hitchhiking with nearby positively selected alleles.

## Discussion

Chickens are among the few domestic species with their progenitors extant in the wild today, providing an excellent system to address questions about evolutionary changes under domestication. In this work, we conducted systematic genomic studies of the domestication history and landscape of hSNPs using the largest genomic dataset from a worldwide sampling of indigenous chickens and their wild counterparts. Our analyses suggest that domestic chickens share a nearly identical demographic history with their direct wild ancestor, the GGS, before the Holocene. Around 12 kya, domestic chickens and GGS have diverged from each other, which is generally consistent with that assessed previously using MSMC based on 50% relative cross-coalescence rate (~9500 ± 3300 years ago) [27]. Subsequently, since domestication, domestic chickens experienced a decline in Ne that was 2.6× more severe than GGS, followed by a recovery towards population expansion. Yet, GGS evolved in a relatively different pattern within this period. We note that the divergence time between domestic chickens and GGS estimated using genomic data is slightly older than the archaeological recordings of chicken domestication; however, the estimated timing of the Ne reduction in domestic chickens generally corresponds with the hypothesized time frame of their domestication [8, 63, 64]. Interestingly, the patterns of Ne differ between domestic chickens and GGS since their split is similar to those observed in both African rice [65] and grape [61] from their wild progenitors, suggesting similar patterns of population size reduction and post-domestication introgression when domesticates spread into new ranges. Our results reveal that the chicken domestication process also followed the same model as what is observed in domestic animals like dogs [66] and horses [12], being initiated with a bottleneck and tailed by a recovery towards population expansion.

Our work reveals that domestic chickens carried 2.95% more hSNPs than GGS, a value comparable to those identified in dogs (2.6%) [13] and rice (~3–4%) [14], but less than that in grape (~5.2%) [61]. Domestic chickens also held a higher ratio of deleterious to synonymous SNPs and a higher frequency of hSNPs compared with GGS. It is possible that the profound historical Ne decline present in all *Gallus gallus* beginning ~80 kya may have induced the accumulation of hSNPs across all lineages at a comparable magnitude, but the domestication bottleneck and/or recent selection for genetic improvement dramatically increased the genetic load of domestic chickens. A similar pattern has already been observed in horses [67] and dogs [13]. Our results support the "cost

of domestication" hypothesis but challenge the proposal of no genetic load during chicken domestication [8, 19, 25, 68]; nonetheless, we could not completely rule out the potential effects from recent genetic improvement and introgression with other jungle fowls on this pattern.

We show that heterozygous hSNPs are accumulated more frequently in the genomes of domestic chickens than GGS, while homozygous ones display a contrary pattern. This is not unexpected, because most harmful mutations are at least partially recessive and therefore could only expose their damaging effects in homozygous states [3]. Especially during breeding practices, harmful mutations in homozygous states are easily observed phenotypically, which promotes purging and breeding decisions, whereas such damaging alleles are masked in heterozygous states and thereby their transmission and accumulation would be facilitated. These results reveal the limitation of current breeding programs in effectively removing potentially damaging effects from the hSNPs while pursuing desirable economic traits. Our study highlights the importance of utilizing genomic information to safeguard genetic improvement through minimizing potential damaging mutations while effectively and sustainably utilizing this species for the poultry industry and biomedical research.

There are several potential caveats in this study. First, our 1K CGP initially aimed to infer the domestication history and evolution of chicken; there is a sampling bias in our study. Our sampling efforts initially focused on diverse and village chickens (which likely present as more "ancient" populations) from Asia (where RJF inhabited) and adjacent regions, while lacking samples from Africa, Oceania, and South America. Even though domestication bottleneck and increased hSNPs are observed in several chicken populations compared with their wild relatives, our samples cannot present the whole genetic diversity of all chickens across the world, and issues on the strength of bottleneck and the number and frequency of these hSNPs accumulated during the early domestication or recent genetic improvement of specific breeds could not be resolved based on our data. Also, there is pervasive gene flow between chickens and other jungle fowls, which also likely result in bias in our estimation of genetic load. Therefore, the magnitudes of bottleneck and genetic load underlying chicken domestication remain open; our analysis provides a result for further testing. Future work by exploring genomes from more heritage chicken lineages and breeds across the world and ancient samples spanning a wide range of periods are necessary to address these questions [67, 69–71]. Second, *TSHR*-559Arg homozygous mice displayed a significant difference in metabolism and development compared with the wild-type mice,

suggesting that this mutation is biologically functional. This supports the early study that investigated the function of this mutation based on birds intercrossed between the ancestral RJF (wild type) and White Leghorn [72]. Because of the profound divergence and potential genetic background difference between chickens and mice, we are not able to directly link any phenotypic changes in mice carrying the chicken allele to domestic chickens and RJFs. Our transgenic experiment provides a preliminary biological indicator to unlock *TSHR* function; however, whether it follows the same biological process in both chickens and mice remains unjustified. Third, despite the pattern of hSNPs in domestic chickens is generally consistent with the observations in other species, our genome coverages are relatively low for both domestic chickens and RJFs, and the levels of heterozygotes and het-hSNPs are likely underestimated. Further validation using higher coverage genomes is warranted. Lastly, our analysis focused on variants in the coding regions; however, non-coding regions are increasingly known to play important regulatory roles, and some variants within these regions likely have significant biological functions [73]; future studies should be designed to explore the evolutionary and functional roles of variants within regulatory regions in domestication and genetic improvement of chickens.

## Conclusions

In conclusion, we systematically characterize the existence of a pre-domestication loss of genetic diversity followed by a domestication bottleneck in chickens, leading to the prominence of high-impact alleles across domestic chicken genomes. Through functional trait analyses, we suggest that these high-impact alleles affect behavior, development, and morphology, and our findings indicate that these alleles are partially under artificial selection pressure while the frequencies of detrimental variants are increased due to drift. This study presents a new page in chicken genomics, calling for a sharpened focus on the comparative genomic diversity of specific breeds and wild lineages, and for intensive functional analyses of high-impact alleles, to understand which contribute to domestication and genetic improvement of particular traits and which are maladaptive. This would enable the development of reliable markers for monitoring the concrete impact of genetic improvement and the purging of deleterious mutations on chicken genome evolution. In addition, our dating of the bottleneck and recovery processes in one of the most heavily relied upon domesticated species in the world has broad implications for understanding the biocultural interactions, translocation, and domestication practices affecting suites of species in Eurasia that were exploited in the past. Our study provides a possibility for further

investigation using breeding experiments and a larger scale of genomes covering a wider sampling of global chickens and fossils.

## Methods
### Genomic datasets
In our 1K Chicken Genome Project (CGP), we leveraged the Illumina sequencing platform and generated 787 genomes from indigenous chickens and jungle fowls [27]. These samples included domestic chickens (*n* = 620) and all five red jungle fowl subspecies (*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), as well as green jungle fowls (*G. varius; n* = 12), Ceylon jungle fowls (*G. lafayettei; n* = 4), and gray jungle fowls (*G. sonnerati; n* = 2). Specifically, *G. g. spadiceus* was sampled from Thailand and Yunnan province of China, and domestic chickens were sampled from villages in Indonesia, Thailand, Vietnam, China, India, Sri Lanka, Bangladesh, Pakistan, Iran, Afghanistan, and Europe. By combining an additional 76 published genomes [44, 56, 57, 74–77] and applying the standard BWA-GATK pipeline [78, 79], 33.4 M SNPs were successfully genotyped for 863 birds, of which ~25 million SNPs were identified in domestic chickens (*n* = 696) and *G. g. spadiceus* (*n* = 45). Genotypes for 9 *G. g. spadiceus* samples (ypt2887–ypt2895) from Thailand and 1 *G. g. spadiceus* sample from Daweishan (ypt570) were admixed with chicken [27] and were removed. This resulting dataset (including 696 domestic chickens and 35 *G. g. spadiceus* samples) was used to perform the genetic diversity and genetic load analyses. The dataset is available at ChickenSD (http://bigd.big.ac.cn/chickensd/; released).

Pair-wise sequential Markovian coalescent (PSMC) [38] and multiple sequential Markovian coalescent (MSMC) analyses require high-coverage genomes for the successful calling of genome-wide heterozygosity. We selected high-coverage genomes from previous studies (find detail in Additional file 1: Table S1) [35, 40, 41, 74], including genomes for GGS and 18 diverse chicken populations (Yunnan chicken, Yunnan game fowl, Emei chicken, Muchuan chicken, Hetian chicken, Tulufan chicken, Lindian chicken, Liyang chicken, Xianju chicken, Baier Yellow chicken, Yunyang Da chicken, Laos chicken, Sri Lankan chicken, Ethiopian chicken, and four European commercial chicken breeds (White Recessive Rocks, Cobb chicken, White Leghorn, and Rhode Island Red)). All reads from these samples were mapped to the chicken reference genome (GRCg6a: https://www.ncbi.nlm.nih.gov/assembly/?term=GCA_000002315.5) using the standard BWA-GATK pipeline [78, 79]. Sequencing coverages for these genomes were calculated using samtools with the "depth" function [80].

### Demographic history inferences
PSMC [38], MSMC [44], and SMC++ [45] were used to estimate the effective population size (Ne) changes in domestic chickens and GGS in the past. For PSMC analysis, consensus sequences of each of the individuals were called using samtools with the "mpileup" command (version: 1.3.1; http://samtools.sourceforge.net/). The loci with less than 1/3 or more than 2 times average read depths were deleted, and sites with consensus qualities below 20 were also removed. PSMC was running with parameters set as "psmc -N25 -t15 -r5 -p 4+25*2+4+6". Input data for MSMC was prepared using the tool generate_multihetsep.py suggested by the author from https://github.com/stschiff/msmc-tools. The genotypes for all samples were phased jointly using Beagle V4.1 [81] with default parameters. For each group, two individuals (four haplotypes) were analyzed.

For running SMC++ (v1.15.2), we sequenced genomes for each of the three GGS samples (IDs ypt3001, ypt3006, and ypt3009) reported [27] previously to coverage over 20-folds. We performed analysis for the regions with reads mapped uniquely that were generated using the SNPable toolkit (http://lh3lh3.users.sourceforge.net/snpable.shtml) with settings "- k=35 and r=0.9". To maximize more populations to be analyzed, five genomes for each population with coverage over 15-folds were used. Input file for SMC++ was generated using the pipeline as the author suggested (https://github.com/popgenmethods/smcpp). Except that Laos chicken, Emei chicken, and Muchuan chicken have less than five genomes with sequencing depth over 15-folds, we analyzed all populations that were analyzed by MSMC and PSMC above. We further included Jingyang chicken and Pengxia chicken for the SMC++ analysis. SMC++ was ran using default parameters.

Finally, we investigated the population histories by analyzing the joint allele frequency spectra using diffusion approximation for demographic inference (*dadi*) [48]. Because we were mostly interested in the joint demographic history of domestic chickens and GGS, we selected a total of 40 genomes (20 for each group; IDs for GGS: 18833, 19912, Ypt570, Xcelris_174, ypt3003, ypt2893_L3_I025, Xcelris_176, ypt3047, ypt2895_L3_I026, ypt2889, ypt3008, ypt3051, ypt3007, ypt3038, ypt3069, ypt3006, ypt3042, ypt3002, ypt2894, and ypt2887; IDs for domestic chickens: YPt648, Ypt638, 43S, Ypt646, ypt3180, ypt2656, 95S_L8_I025, Ypt606, 19S_L4_I026, 87S_L5_I010, ypt948_L2_I034, Ypt645, 77S_L4_I029, ypt910_L3_I005, 88S_L4_I011, 44S_L4_I001, 130S_L4_I044, ypt907_L6_I002, 39S_L6_I053, and 21S_L6_I028) from the 1K CGP. To avoid evolutionary restrictions as much as possible, we excluded coding regions. We also masked the repeated (annotations from NCBI: https://www.ncbi.nlm.nih.gov/) and low-complexity regions identified using mdust [78]. Finally, 47,307 autosomal

Wang *et al. BMC Biology*    (2021) 19:118

Page 12 of 16

regions of at least 1 kb spanning a total of 70,919,324 bp were used for the demographic analysis. We computed two-dimensional site frequency spectra using ANGSD [82], as described previously [70]. We examined four demographic models (Additional file 1: Figure S3): (A) constant without gene flow, (B) constant with asymmetric gene flow, (C) constant-growth/reduction with asymmetric migrations, and (D) constant-growth/reduction with asymmetric migrations. For each model, we ran three sets of increasingly focused optimizations before performing the final model selection. Models were compared using the Akaike Information Criterion (AIC), and the replicate with the highest likelihood for each model was used to calculate AIC and deltAIC. To calculate the confidence interval for the parameters in our best-fitting model, we applied non-parametric bootstrapping (100 replicates).

Estimations from PSMC, MSMC, SMC++, and *dadi* were scaled using a generation time (g) of 1 year and a mutation rate (μ) of $1.91 \times 10^{-9}$ substitutions per site per year [42]. We used VCFtools [83] to average the population-based nucleotide diversity (π) [84] in domestic chickens and GGS ("--window-pi 50000 --window-pi-step 25000").

### Identification of high-impact mutations and assessment of genetic loads

To examine the evolution of hSNPs in chickens before and after their domestication, we followed a similar pipeline as described previously [60, 61] to analyze the 33.4 M SNPs called from the 863 genomes. First, we retrieved non-synonymous mutations as annotated by ANNOVAR [85] and searched the chicken genome annotations from the ENSEMBL database (version 83: http://dec2015.archive.ensembl.org/index.html). The ancestral state of a variant was inferred from the green jungle fowl. To predict the effect of a missense mutation on a protein, we applied the PROVEAN software [50] in searching the non-redundant protein database (download from NCBI: https://www.ncbi.nlm.nih.gov/). The prediction was based on evolutionary conservation by comparing the query and target sequences. Similar to the early study [14], a mutation with a PROVEAN score less than –2.5 was considered to be harmful, and such kinds of variants were labeled as hSNPs.

To obtain a global perspective on the functions of genes carrying such hSNPs, we used g:Profiler [86] to retrieve the functional enrichment terms, including Gene Ontology (GO, KEGG pathways) and Human Phenotype Ontologies (HPOs). To assess the landscape of genetic loads over chicken domestication, we calculated the number and frequency of hSNPs per individual or region and compared them with those of synonymous mutations.

Furthermore, we retrieved genomic regions of putatively selective sweeps that were measured by locus-specific

branch length (LSBL) statistics [87] in the combination of LSBL1 (chicken; *G. g. spadiceus*, *G. g. jabouillei*) and LSBL2 (chicken; *G. g. spadiceus*, *G. g. murghi*), as well as π-ratio ($\pi_{G.g.spadiceus}/\pi_{chicken}$) [84] from our previous study [27]. For sweep regions identified by each of the three statistics, we compared the number and frequency of hSNPs within the sweeps to the remaining genomic regions between domestic chickens and GGS.

### Construction of *TSHR* knock-in mice

The allele *TSHR*-Gly558 (chr5:40,089,599G) in domestic chickens is highly conserved across vertebrates and corresponds to the mice-*TSHR*-Gly559 (c.1675G) in the 10th exon of transcript Tshr-202 (http://www.ensembl.org/Mus_musculus/Transcript/Exons?db=core;g=ENSMUSG00000020963;r=12:91400994-91540509;t=ENSMUST00000021346). A C57BL/6 mouse model with a mutation at the mouse *TSHR* locus (p. Gly559Arg; c.1675G>A) was constructed by CRISPR/Cas-mediated genome engineering (Shanghai Biomodel Organism Science & Technology Development Co., Ltd). Briefly, Cas9 mRNA, gRNA, and donor DNA were micro-injected into the fertilized eggs of C57BL/6J mice to obtain F0 generation mice with the mutation of the target site (Additional file 1: Figure S5 and Table S6). The F0 generation mice were mated with C57BL/6J mice to obtain positive and homozygous F1 generation mice. All mice had free access to food and water. All experiments were performed following the Health Guide for the Care and Use of Laboratory Animals and were approved by the Ethics Committee of the Kunming Institute of Zoology, CAS.

### Metabolism assay of the transgenic mice

A Comprehensive Laboratory Animal Monitoring System (CLAMS) was used to monitor and analyze the indexes of metabolism and feed intake of the transgenic mice. Eight-week-old male mice were weighed (n = 8*2) and placed in CLAMS (PRO-MRR-8) to measure the metabolism of homozygous (HO) and wild-type mice at 30 °C, 18 °C, and 5 °C for 72 h. The levels of oxygen consumption ($VO_2$), carbon dioxide exhalation ($VCO_2$), calorie consumption, and water and food intake were recorded every 5 min for each mice (including HO and wild type). All mice had free access to water and food and were subjected to the same day-night cycle during the examination. Statistical significance was measured by Student's test (two-tailed), and $P < 0.05$ was accepted to be significant.

Wang *et al. BMC Biology*        (2021) 19:118

Page 13 of 16

# Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-021-01052-x.

---

**Additional file 1: Figure S1.** Nucleotide diversity for *G. g. spadiceus* and chicken populations (grouped by samling locations and breeds). **Figure S2.** Demographic histories for *G. g. spadiceus* and diverse chicken groups by PSMC. A total of 18 chicken populations were included in this analysis. **Figure S3.** Four tested demographic models for *dadi* analysis. Nanc, ancestral population size before the split; T, timepoints; m, migrations; Napop, ancestral population size after split. Ncpop, current population size. Arrows depict migration directions. **Figure S4.** Comparing observed data and model allele frequency spectrum for the best model (Model 3). **Figure S5.** Provean-scores for nonsynonymous mutations (for all mutations, left; and for mutations with Provean-scores ≤ −2.5, right) in microchromosomes, macrochromosomes, and intermediate chromosomes. **Figure S6.** Pipeline for constructing the mouse model with a mutation at the mouse *TSHR* locus (p. Gly559Arg; c.1675G>A). **Figure S7.** Photograph showing *TSHR*-559Arg knock-in homozygous (HO) and wild-type mice at 10 months old. **Figure S8.** Gly558Arg knock-in mice consumed less food than wild-type. *, $P < 0.05$. Statistical significance was measured by the Student's t test. N = 8 for both HO and wild-type male mice were used in each test. **Figure S9.** Number and ratio of high-impact mutations among chicken populations. GGS, *G. g. spadiceus*; DC, all domestic chickens; WL, White Leghorn; TC, Tibetan chicken, XJ, Xinjiang local chicken; You; Beijing You chicken. **Figure S10.** Number and frequency of deleterious mutations in the genomic regions of putatively selective sweeps. **Table S1.** Information for high-coverage genomes used for PSMC, MSMC, and SMC++ analyses. **Table S2.** Estimations of likelihoods and AIC scores from four demographic models. **Table S3.** Summary of population histories calculated from 2D–SFS. Confidence intervals (95%) were obtained by bootstrapping all sites and performing parameter inference on each bootstrap dataset with 100 runs. **Table S4.** Distribution of variants identified in dog, sheep, goat, cattle, pig, and horse. This data is from our previous publication. **Table S5.** GO enrichment for genes carrying nonsynonymous mutation with provean-score of <−10. **Table S6.** Guide RNA sequences for the exon 10 of mouse-*TSHR*.

---

## Authors' contributions

Y.-P.Z., J.-L.H., D.-D.W., M.-S.W., and Y.J. conceived the project and designed the research. M.-S.W., X.G., H.A., Z.-Q.Z., S.W., L. Z., S.-H.F., and M.L. performed the analysis. J.-J.Z. and M.-M.Y. conducted the wet lab experiments. X.-M.L., M.K., J.-L.H., A.E., K.-X.Q., C.S., N.Y.H., H.A., S.S., M.S.A.Z., S.K., S. S., H.K.-K., E.L., S.C., H.G.T.N.G., T.M.S., H.Z., A.K.F.H. B., M.S.K., G.L.L.P.S., T.-T. Y., Y.-M.W., O.A.M., and M.N.M.I. did the sample collection, prepared the DNA samples, and contributed genome sequencing data for 1K CGP. M.-S.W., J.-L.H., H.A., M.B., and R.M. drafted the manuscript with input from all authors. J.-L.H., M.B., R.M., Y.-P.Z., D.-D.W., M.-S.W., O.H., and A.E. revised the manuscripts. All authors read and approved the final manuscript.

## Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. Information for the published dataset for estimating demographic histories is available in Additional file 1: Table S1. The genotype data we analyzed (in VCF format) and three GGS genomes we generated were available at ChickenSD (http://bigd.big.ac.cn/chickensd/).

## Declarations

### Ethics approval and consent to participate
Not applicable.

### Consent for publication
Not applicable.

### Competing interests
The authors declare that they have no competing interests.

### Author details
[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 650223, China. [2]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China. [3]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [4]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA. [5]College of Animal Science and Technology, Anhui Agricultural University, Hefei 230036, China. [6]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China. [7]Museum Zoologicum Bogoriense, Research Center for Biology, Indonesian Institute of Science (LIPI), Cibinong, Bogor 16911, Indonesia. [8]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China. [9]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, The Cooperative Innovation Center for Sustainable Pig Production, Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China. [10]Zoological Survey of India, New Alipore, Kolkata, West Bengal 700053, India. [11]School of Agriculture and Natural Resources, University of Phayao, Phayao 56000, Thailand. [12]Unit of Excellence on Biodiversity and Natural Resources Management, University of Phayao, Phayao 56000, Thailand. [13]Department of Animal Science, Shahid Bahonar University of Kerman, P.O. Box 76169133, Kerman, Iran. [14]Faculty of Agriculture, University of Ruhuna, Matara, Sri Lanka. [15]Institute of Animal Husbandry, Biotechnology and Nature Conservation, University of Debrecen, Debrecen H-4032, Hungary. [16]Institute of Biotechnology, School of Agriculture, Shiraz University, P.O. Box 1585, Shiraz, Iran. [17]Center for Neurobiology and Brain Restoration, Skolkovo Institute of Science and Technology, Moscow 143026, Russia. [18]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650204, China. [19]Dipartimento di Scienze Agrarie, Alimentarie Ambientali, University of Perugia, 06123 Perugia, Italy. [20]BGI-Shenzhen, Beishan Industrial Zone, Shenzhen 518083, China. [21]Laboratory of Animal Genetics, Breeding and Reproduction, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Ministry of Agriculture of China, Beijing 100193, China. [22]Bangladesh Agricultural University, Mymensingh 2202, Bangladesh. [23]Cholistan University of Veterinary and Animal Sciences, Bahawalpur, Pakistan. [24]Department of Animal Science, University of Peradeniya, Peradeniya 20400, Sri Lanka. [25]National Institute of Animal Husbandry, Hanoi, Vietnam. [26]Livestock Genetics Program, International Livestock Research Institute (ILRI), Nairobi 00100, Kenya. [27]China National Genebank, BGI-Shenzhen, Shenzhen 518083, China. [28]Centre for Social Evolution, Department of Biology, University of Copenhagen, DK-1870 Copenhagen, Denmark. [29]Yunnan Academy of Grassland and Animal Science, Kunming 650212, China. [30]Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham,

Wang *et al. BMC Biology* (2021) 19:118

Page 14 of 16

Nottingham NG7 2RD, UK. [31]Livestock Genetics Program, International Livestock Research Institute (ILRI), P.O. Box 5689, Addis Ababa, Ethiopia. [32]Wageningen University & Research - Animal Breeding and Genomics, 6708 PB Wageningen, The Netherlands. [33]State Key Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China.

## References

1. Bosse M, Megens HJ, Derks MFL, de Cara AMR, Groenen MAM. Deleterious alleles in the context of domestication, inbreeding, and selection. Evol Appl. 2019;12(1):6–17 https://doi.org/10.1111/eva.12691.
2. Makino T, Rubin CJ, Carneiro M, Axelsson E, Andersson L, Webster MT. Elevated proportions of deleterious genetic variation in domestic animals and plants. Genome Biol Evol. 2018;10(1):276–90 https://doi.org/10.1093/gbe/evy004.
3. Agrawal AF, Whitlock MC. Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics. 2011;187(2):553–66 https://doi.org/10.1534/genetics.110.124560.
4. Groß C, Bortoluzzi C, de Ridder D, Megens H-J, Groenen MAM, Reinders M, et al. Evolutionarily conserved non-protein-coding regions in the chicken genome harbor functionally important variation. bioRxiv. 2020; https://doi.org/10.1101/2020.03.27.012005.
5. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alfoldi J, Martinez Barrio A, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science (New York, NY). 2014;345(6200):1074–9 https://doi.org/10.1126/science.1253714.
6. Gerbault P, Allaby RG, Boivin N, Rudzinski A, Grimaldi IM, Pires JC, et al. Storytelling and story testing in domestication. Proc Natl Acad Sci U S A. 2014;111(17):6159–64 https://doi.org/10.1073/pnas.1400425111.
7. Frantz LAF, Bradley DG, Larson G, Orlando L. Animal domestication in the era of ancient genomics. Nat Rev Genet. 2020;21(8):449–60. https://doi.org/10.1038/s41576-020-0225-0.
8. Larson G, Fuller DQ. The evolution of animal domestication. Annu. Rev. Ecol. Evol. Syst. 2014;45(1):115–36 https://doi.org/10.1146/annurev-ecolsys-110512-135813.
9. Gaut BS, Seymour DK, Liu Q, Zhou Y. Demography and its effects on genomic variation in crop domestication. Nat Plants. 2018;4(8):512–20 https://doi.org/10.1038/s41477-018-0210-1.
10. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, et al. Convergent genomic signatures of domestication in sheep and goats. Nat Commun. 2018;9(1):813 https://doi.org/10.1038/s41467-018-03206-y.
11. Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. Nat Rev Genet. 2015;16(6):333–43 https://doi.org/10.1038/nrg3931.
12. Schubert M, Jonsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc Natl Acad Sci U S A. 2014;111(52):E5661–9 https://doi.org/10.1073/pnas.1416991111.
13. Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vila C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci U S A. 2016;113(1):152–7 https://doi.org/10.1073/pnas.1512501113.
14. Liu Q, Zhou Y, Morrell PL, Gaut BS. Deleterious variants in Asian rice and the potential cost of domestication. Mol Biol Evol. 2017;34(4):908–24 https://doi.org/10.1093/molbev/msw296.
15. Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, et al. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (Oryza glaberrima). Mol Ecol. 2014;23(9):2210–27 https://doi.org/10.1111/mec.12738.
16. Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci U S A. 2013;110(28):E2655–62 https://doi.org/10.1073/pnas.1309606110.
17. Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends Genet. 2006;22(3):126–31 https://doi.org/10.1016/j.tig.2006.01.004.
18. Brown TA. Is the domestication bottleneck a myth? Nat Plants. 2019;5(4):337–8 https://doi.org/10.1038/s41477-019-0404-1.
19. Frantz LA, Larson G. A genetic perspective on the domestication continuum. In: Hybrid communities: biosocial approaches to domestication and other trans-species relationships: Taylor & Francis; 2018. http://qmro.qmul.ac.uk/xmlui/handle/123456789/53406.
20. Frantz LA, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat Genet. 2015;47(10):1141–8 https://doi.org/10.1038/ng.3394.
21. Wallberg A, Han F, Wellhagen G, Dahle B, Kawata M, Haddad N, et al. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee Apis mellifera. Nat Genet. 2014;46(10):1081–8 https://doi.org/10.1038/ng.3077.
22. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464(7288):587–91 https://doi.org/10.1038/nature08832.
23. Qanbari S, Rubin CJ, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation in modern chicken. PLoS Genet. 2019;15(4):e1007989 https://doi.org/10.1371/journal.pgen.1007989.
24. Bortoluzzi C, Bosse M, Derks MFL, Crooijmans R, Groenen MAM, Megens HJ. The type of bottleneck matters: insights into the deleterious variation landscape of small managed populations. Evol Appl. 2020;13(2):330–41 https://doi.org/10.1111/eva.12872.
25. Larson G, Burger J. A population genetics view of animal domestication. Trends Genet. 2013;29(4):197–205 https://doi.org/10.1016/j.tig.2013.01.003.
26. Guo X, He XX, Chen H, Wang ZC, Li HF, Wang JX, et al. Revisiting the evolutionary history of domestic and wild ducks based on genomic analyses. Zool Res. 2021;42(1):43–50. https://doi.org/10.24272/j.issn.2095-8137.2020.133.
27. Wang MS, Thakur M, Peng MS, Jiang Y, Frantz LAF, Li M, et al. 863 genomes reveal the origin and domestication of chicken. Cell Res. 2020;30(8):693–701 https://doi.org/10.1038/s41422-020-0349-y.
28. Fan Z, Silva P, Gronau I, Wang S, Armero AS, Schweizer RM, et al. Worldwide patterns of genomic variation and admixture in gray wolves. Genome Res. 2016;26(2):163–73. https://doi.org/10.1101/gr.197517.115.
29. Zhang Z, Khederzadeh S, Li Y. Deciphering the puzzles of dog domestication. Zool Res. 2020;41(2):97–104. https://doi.org/10.24272/j.issn.2095-8137.2020.002.
30. Zeder MA. The origins of agriculture in the Near East. Curr Anthropol. 2011;52(S4):S221–35 https://doi.org/10.1086/659307.
31. Zhang Y, Colli L, Barker JSF. Asian water buffalo: domestication, history and genetics. Anim Genet. 2020;51(2):177–91 https://doi.org/10.1111/age.12911.
32. Guo X, Wang ZC, Wang S, Li HF, Suwannapoom C, Wang JX, et al. Genetic signature of hybridization between Chinese spot-billed ducks and domesticated ducks. Anim Genet. 2020;51(6):866–75 https://doi.org/10.1111/age.13002.
33. Wang MS, Wang S, Li Y, Jhala Y, Thakur M, Otecko NO, et al. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Mol Biol Evol. 2020;37(9):2616–29 https://doi.org/10.1093/molbev/msaa113.
34. Lawler A. Why did the chicken cross the world? New York: Atria Books; 2014.
35. Lawal RA, Martin SH, Vanmechelen K, Vereijken A, Silva P, Al-Atiyat RM, et al. The wild species genome ancestry of domestic chickens. BMC Biol. 2020;18(1):13 https://doi.org/10.1186/s12915-020-0738-1.
36. Collias NE, Saichuae P. Ecology of the red jungle fowl in Thailand and Malaya with reference to the origin of domestication. Nat Hist Bull Siam Soc. 1967;22:189–209.
37. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491(7424):393–8 https://doi.org/10.1038/nature11622.
38. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475(7357):493–6 https://doi.org/10.1038/nature10231.
39. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol Ecol. 2016;25(5):1058–72 https://doi.org/10.1111/mec.13540.
40. Luo W, Luo C, Wang M, Guo L, Chen X, Li Z, et al. Genome diversity of Chinese indigenous chicken and the selective signatures in Chinese

Wang *et al. BMC Biology* (2021) 19:118

Page 15 of 16

gamecock chicken. Sci Rep. 2020;10(1):14532. https://doi.org/10.1038/s41598-020-71421-z.

41. Li D, Li Y, Li M, Che T, Tian S, Chen B, et al. Population genomics identifies patterns of genetic diversity and selection in chicken. BMC Genomics. 2019; 20(1):263 https://doi.org/10.1186/s12864-019-5622-4.

42. Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JB, Backström N, et al. Molecular evolution of genes in avian genomes. Genome Biol. 2010;11(6): R68. https://doi.org/10.1186/gb-2010-11-6-r68.

43. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr Biol. 2015;25(10):1375–80 https://doi.org/10.1016/j.cub.2015.03.047.

44. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014;46(8):919–25 https://doi.org/10.1038/ng.3015.

45. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 2017;49(2): 303–9 https://doi.org/10.1038/ng.3748.

46. Bosse M, Megens HJ, Madsen O, Frantz LA, Paudel Y, Crooijmans RP, et al. Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations. Mol Ecol. 2014;23(16):4089–102 https://doi.org/10.1111/mec.12807.

47. Orozco-terWengel PA, Bruford MW. Mixed signals from hybrid genomes. Mol Ecol. 2014;23(16):3941–3 https://doi.org/10.1111/mec.12863.

48. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009;5(10):e1000695 https://doi.org/10.1371/journal.pgen.1000695.

49. Lohmueller KE, Bustamante CD, Clark AG. The effect of recent admixture on inference of ancient human population history. Genetics. 2010;185(2):611–22 https://doi.org/10.1534/genetics.109.113761.

50. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PloS One. 2012;7(10):e46688 https://doi.org/10.1371/journal.pone.0046688.

51. Axelsson E, Webster MT, Smith NG, Burt DW, Ellegren H. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res. 2005;15(1): 120–5. https://doi.org/10.1101/gr.3021305.

52. Meseret S. A review of poultry welfare in conventional production system. Livest. Res. Rural. Dev. 2016;28:12.

53. Hirsch V. Legal protections of the domestic chicken in the United States and Europe. East Lansing: Michigan State University College of Law; 2003.

54. Loog L, Thomas MG, Barnett R, Allen R, Sykes N, Paxinos PD, 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. 2017;34(8):1981–90 https://doi.org/10.1093/molbev/msx142.

55. Pan S, Lin Y, Liu Q, Duan J, Lin Z, Wang Y, et al. Convergent genomic signatures of flight loss in birds suggest a switch of main fuel. Nat Commun. 2019;10(1):2756 https://doi.org/10.1038/s41467-019-10682-3.

56. Wang MS, Zhang RW, Su LY, Li Y, Peng MS, Liu HQ, et al. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. Cell Res. 2016;26(5):556–73 https://doi.org/10.1038/cr.2016.44.

57. Wang MS, Huo YX, Li Y, Otecko NO, Su LY, Xu HB, et al. Comparative population genomics reveals genetic basis underlying body size of domestic chickens. J Mol Cell Biol. 2016;8(6):542–52 https://doi.org/10.1093/jmcb/mjw044.

58. Abe E, Marians RC, Yu W, Wu XB, Ando T, Li Y, et al. TSH is a negative regulator of skeletal remodeling. Cell. 2003;115(2):151–62 https://doi.org/10.1016/S0092-8674(03)00771-2.

59. Grottesi A, Gabbianelli F, Valentini A, Chillemi G. Structural and dynamic analysis of G558R mutation in chicken TSHR gene shows altered signal transduction and corroborates its role as a domestication gene. Anim Genet. 2020;51(1):51–7 https://doi.org/10.1111/age.12880.

60. Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, et al. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. Nat Genet. 2017;49(6):959–63 https://doi.org/10.1038/ng.3845.

61. Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. Proc Natl Acad Sci U S A. 2017;114(44):11715–20 https://doi.org/10.1073/pnas.1709257114.

62. He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, et al. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nat Genet. 2019;51(5):896–904 https://doi.org/10.1038/s41588-019-0382-2.

63. Peters J, Lebrasseur O, Best J, Miller H, Fothergill T, Dobney K, et al. Questioning new answers regarding Holocene chicken domestication in China. Proc Natl Acad Sci U S A. 2015;112(19):E2415 https://doi.org/10.1073/pnas.1503579112.

64. Tixier-Boichard M, Bed'hom B, Rognon X. Chicken domestication: from archeology to genomics. C R Biol. 2011;334(3):197–204 https://doi.org/10.1016/j.crvi.2010.12.012.

65. Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, et al. Domestication history and geographical adaptation inferred from a SNP map of African rice. Nat Genet. 2016;48(9):1083–8 https://doi.org/10.1038/ng.3633.

66. Wang GD, Zhai W, Yang HC, Wang L, Zhong L, Liu YH, et al. Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res. 2016;26(1):21–33 https://doi.org/10.1038/cr.2015.147.

67. Fages A, Hanghoj K, Khan N, Gaunitz C, Seguin-Orlando A, Leonardi M, et al. Tracking five millennia of horse management with extensive ancient genome Time Series. Cell. 2019;35(7):473–5 https://doi.org/10.1016/j.cell.2019.03.049.

68. Bosse M. No "doom" in chicken domestication? PLoS Genet. 2019;15(5): e1008089 https://doi.org/10.1371/journal.pgen.1008089.

69. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, et al. Ancient genomic changes associated with domestication of the horse. Science (New York, NY). 2017;356(6336):442–5 https://doi.org/10.1126/science.aam5298.

70. Zheng Z, Wang X, Li M, Li Y, Yang Z, Wang X, et al. The origin of domestication genes in goats. Sci Adv. 2020;6(21):eaaz5216 https://doi.org/10.1126/sciadv.aaz5216.

71. Allaby RG, Ware RL, Kistler L. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. Evol Appl. 2019;12(1):29–37 https://doi.org/10.1111/eva.12680.

72. Karlsson AC, Fallahshahroudi A, Johnsen H, Hagenblad J, Wright D, Andersson L, et al. A domestication related mutation in the thyroid stimulating hormone receptor gene (TSHR) modulates photoperiodic response and reproduction in chickens. Gen Comp Endocrinol. 2016;228:69–78 https://doi.org/10.1016/j.ygcen.2016.02.010.

73. Liu S, Yu Y, Zhang S, Cole JB, Tenesa A, Wang T, et al. Epigenomics and genotype-phenotype association analyses reveal conserved genetic architecture of complex traits in cattle and human. BMC Biol. 2020;18(1):80 https://doi.org/10.1186/s12915-020-00792-6.

74. Wang MS, Li Y, Peng MS, Zhong L, Wang ZJ, Li QY, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol Biol Evol. 2015;32(7):1880–9 https://doi.org/10.1093/molbev/msv071.

75. Yi G, Qu L, Liu J, Yan Y, Xu G, Yang N. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. BMC Genomics. 2014;15(1):962. https://doi.org/10.1186/1471-2164-15-962.

76. Fan WL, Ng CS, Chen CF, Lu MY, Chen YH, Liu CJ, et al. Genome-wide patterns of genetic variation in two domestic chickens. Genome Biol Evol. 2013;5(7):1376–92 https://doi.org/10.1093/gbe/evt097.

77. Wang MS, Otecko NO, Wang S, Wu DD, Yang MM, Xu YL, et al. An evolutionary genomic perspective on the breeding of dwarf chickens. Mol Biol Evol. 2017;34(12):3081–8 https://doi.org/10.1093/molbev/msx227.

78. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics (Oxford, England). 2014;30(20):2843–51 https://doi.org/10.1093/bioinformatics/btu356.

79. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303. https://doi.org/10.1101/gr.107524.110.

80. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078–9 https://doi.org/10.1093/bioinformatics/btp352.

81. Browning BL, Browning SR. Genotype imputation with millions of reference samples. Am J Hum Genet. 2016;98(1):116–26 https://doi.org/10.1016/j.ajhg.2015.11.020.

82. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics. 2014;15(1):356 https://doi.org/10.1186/s12859-014-0356-4.

83. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics (Oxford, England). 2011; 27(15):2156–8 https://doi.org/10.1093/bioinformatics/btr330.

84. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 1979;76(10):5269–73 https://doi.org/10.1073/pnas.76.10.5269.

85. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38(16):e164 https://doi.org/10.1093/nar/gkq603.

86. Reimand J, Arak T, Vilo J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. 2011;39(suppl_2):W307–15. https://doi.org/10.1093/nar/gkr378.

87. Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, et al. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genomics. 2004;1(4):274–86 https://doi.org/10.1186/1479-7364-1-4-274.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.