Kadaknath chicken specific hitchhiking of *BPIL* with the FM locus

Sagar Sharad Shinde[1], Ashutosh Sharma[1], Nagarjun Vijay[1]

[1]Computational Evolutionary Genomics Lab, Department of Biological Sciences, IISER Bhopal, Bhauri, Madhya Pradesh, India

*Correspondence: nagarjun@iiserb.ac.in

Keywords: Kadaknath, black-bone chicken, introgression, genetic linkage, fibromelanosis

Abstract

Black-bone-chicken (BBC) breeds are famous for their distinctive meat texture and taste. However, the relationship between BBC breeds, especially the Kadaknath breed native to India, is poorly understood. Here, we establish that all BBC breeds, including Kadaknath, share a common complex chromosomal rearrangement at the FM locus on the 20th chromosome, which is responsible for the melanin hyperpigmentation throughout the body. While we find breed-specific gene flow between BBC and commercial breeds, the shared ancestry of all BBC breeds and pattern of isolation by distance suggest Kadaknath is closely related to Tibetan BBC. Interestingly, two FM locus proximal regions (~70Kb and ~300 Kb) have signatures of selection unique to Kadaknath. These regions harbor several genes with protein-coding changes, with the Bactericidal/permeability-increasing-protein-like (*BPIL*) gene having two Kadaknath-specific changes within the protein domains. The absence of these changes in other BBC breeds suggests this variation arose after the separation of these breeds. Our results indicate that the rearrangement at the FM locus and protein-coding changes in the *BPIL* gene were selected simultaneously in Kadaknath due to close physical linkage. The identification of this FM locus proximal selective sweep sheds light on the distinctiveness of Kadaknath compared to other BBC.

Introduction

Domestication of chicken during the Neolithic period involved a complex pattern of interbreeding with various jungle fowl species[1–4]. After domestication, chicken has spread worldwide and occurs as commercial, fancy, and indigenous village breeds. Humans use chicken as a research model due to their physiology and behaviour[5], as game fowl and for religious reasons, or more commonly for egg or meat production[6]. Chicken is the most preferred source of meat for humans due to its easy availability and affordability[7]. Hence, understanding the genetics of meat features is commercially relevant[8]. Although understudied, indigenous village chicken breeds with unique properties provide an opportunity to understand the genetics of meat. For instance, the black-bone chicken (BBC) is a delicacy due to its texture, color, firmness, and flavor[9]. The BBC result from melanin deposition throughout the body, i.e., melanin hyperpigmentation or fibromelanosis caused by the FM allele[10]. Bateson and Punnet were the pioneers in identifying the autosomal dominant FM allele[11]. Modern studies found a chromosomal rearrangement on chromosome 20 is

involved in the FM locus[12,13]. The overexpression of the *EDN3* gene located within the FM locus is responsible for dermal pigment deposition[12,14].

BBC breeds have anti-fatigue and anti-hypoxic abilities, with its meat having antioxidant properties[15,16], high carnosine[17,18], and lower fat and cholesterol content[9,19]. Some BBC breeds also have local adaptations. For instance, the Korean Ogye has improved fetal viability and innate immunity against microbial and viral infections[20]. Nonetheless, BBC is famous for its tender meat, delicious flavor, and use in traditional medicine[16]. BBC breeds occur globally and have distinct names, such as Ayam Cemani (Indonesia), Black H'Mong (Vietnam), Tuzo (Argentina), Svarthöna (Sweden)[13], Yeonsan Ogye (Korea)[21], and Thai BBC (Thailand)[22]. China has a high diversity of BBC breeds, including Silkie, Jiangshan, Lueyang, Sichuan, Xingwen, Yugan, Dehua, Jinhu, Muchuan, Wumeng, Yanjin, Xichuan, Tuanfu, Wuliangshan, Emei and Miyi fowl [23,24,25,26,16]. India has a single breed of BBC commonly known as Kadaknath.

Kadaknath is native to the Jhabua and Dhar districts of Madhya Pradesh, but its farming has recently spread across India. At least three distinctive phenotypes (i.e., Jetblack, Pencil, and Golden) occur within the Kadaknath breed[27]. In Jetblack, all the body parts like plumage, comb, internal organs, eyes, skin, beak, shank, and claw are entirely black, whereas Pencil and Golden have white and golden color patches on the plumage, respectively[27]. Several unique characteristics, such as earlier egg-laying maturity, high protein content, better disease resistance, and adaptation to the local environment, are attributed to Kadaknath chicken[18,28–33]. Understanding the genetics behind these traits will help establish the uniqueness of Kadaknath and guide de-extinction efforts and breeding programs.

Despite its immense popularity and commercial value, the genomics of the Kadaknath chicken breed has received limited attention. Therefore, further research and genomic analysis are required to understand adaptations in this breed and its genetic history. Moreover, the prevalence of interbreeding with commercial chicken lines to increase productivity and a declining population pose a threat of extinction to this native chicken breed. The high levels of interbreeding also present an opportunity to identify genomic regions involved in coding for breed-specific traits. The aims of this study are to (1) Perform whole-genome resequencing of Kadaknath and evaluate its relationship with other black-bone and non-black-bone chicken breeds. (2) Assess whether Kadaknath and other BBC breeds

share a common origin by comparing the chromosomal rearrangement junction at the FM locus. (3) Use population genetic statistics to identify signatures of selection in Kadaknath compared to other BBC breeds, and (4) Identify the genomic patterns of introgression from commercial chicken breeds. Our results offer an example of the co-selection of traits regulating immune response, meat quality, and plumage color providing new approaches to improve breed quality and productivity in chicken simultaneously.

Results

Whole-genome re-sequencing

We generated >25X coverage whole-genome resequencing Illumina data for 12 chicken individuals from India. Our collection includes 9 Kadaknath samples as representatives of all three extant morphs Jetblack (JETB n=3), Golden (GOLD n=3), and Pencil (PENC n=3). We also sampled one individual with a golden-pencil-like phenotype (GOPE n=1), one with complete white plumage with black-bone (CROS n=1), and one non-black-bone individual where plumage was black (NONB) (**Supplementary Fig. 1 and Table 1**). We compared these Kadaknath chicken genomes with public resequencing datasets of BBC breeds. The sampling location (see **Fig. 1A** for black-bone chicken) of individuals analyzed in this study is spread across Asia (for more detail, see **Table 2,** which includes the complete list of all chicken breeds examined). A detailed map of China depicts the locations of all the Chinese black-bone breeds (**Supplementary Fig. 2**).

Kadaknath is a distinct breed of black-bone fowl

The major axes of genetic variation (PC1:12.7% and PC2:7.17%) separated the commercial (RIRL: Rhode Island Red, WHLL: White Leghorn, and CWLH: Commercial White Leghorn) from black-bone and other native indigenous breeds (see **Fig. 1B, 1C** and **Supplementary Fig. 3A**). We observed that all Kadaknath individuals form a single cluster distinct (PC3: 4.71%) from other black-bone (from Indonesia, China, and South Korea) and native chicken breeds (**Fig. 1B**, **1C** and **Supplementary Fig. 3B, 4**). Surprisingly, the broiler (BROC) was genetically similar to Kadaknath and Yeonsan Ogye (YOSK), hinting at introgression from commercial breeds (see **Supplementary Fig. 3A**). In contrast to the lower genetic (π = 0.002, θ = 0.002) diversity in YOSK, Kadaknath and Chinese BBC have comparable (π = 0.004 and 0.003, θ = 0.003 in both) genetic diversity (**Supplementary Fig. 5A, B**). Admixture analysis suggests that all black-bone breeds except Kadaknath and Miyi fowl have a substantial genetic contribution from commercial breeds (see **Fig. 1C** and **Supplementary Fig. 4**).

Moreover, Kadaknath has a sub-structure but remains distinct from the other black-bone breeds (**Supplementary Fig. 6, 7**). In contrast to the nuclear genome, the mitochondrial genome haplotype network did not separate Kadaknath from other BBC breeds (**Supplementary Fig. 7A, 7B**).

To delineate the history of different black-bone breeds, we first evaluated if a pattern of isolation by distance (IBD) is prevalent. The IBD plot shows a consistent increase in genetic distance with an increase in geographic distance (see **Supplementary Fig. 8**, mantels r=0.64, p-value=0.0002). The IBD pattern persisted even after we repeated the analysis using ANGSD/variant-call-based estimates of F_{ST} after excluding various populations to avoid errors due to auto-correlation and biased estimates from isolated populations (see **Supplementary Fig. 8-15**). Among the black-bone breeds, Tibetan black-bone (TBTC) and LCTMJ chickens are genetically closest to Kadaknath (mean F_{ST}~0.17). The other black-bone breeds (LCEM, mean F_{ST}=0.21; XBBC, mean F_{ST}=0.22; LCMY, mean F_{ST}=0.3 and YOSK, mean F_{ST}=0.36) occur at increasing genetic distances from Kadaknath (see **Table 3**). Analysis of our dataset found that Kadaknath has almost twice the number of private alleles than Chinese and Korean BBC breeds, even after excluding alleles found in dbSNP (**Supplementary Fig. 16-17**).

All black-bone chicken breeds share the rearrangement at FM locus

The normalized short-read coverage in black-bone (**Fig. 2A**) abruptly increases at Dup1 and Dup2 regions, while no such increase occurs in non-black-bone (**Fig. 2B**) chicken. The drastic change in coverage at Dup1 and Dup2 boundaries occurs at the same base in all black-bone chickens (see inset in **Fig. 2A**). In the case of non-black-bone chickens, no change in coverage occurs at this position (see inset in **Fig. 2B**). We also identified the black-bone-specific FM-locus junctions START-DUP1-END-END-DUP2-START and END-DUP1-START-START-DUP2-END using published PacBio data (see **Methods**). Short-read coverage spanning these black-bone-specific FM-locus junctions is present in black-bone and missing in non-black-bone chicken (**Supplementary Fig. 18**). The nucleotide differentiation (F_{ST}) between black-bone and non-black-bone chicken along chromosome 20 is elevated at the Dup1 and Dup2 regions (see **Fig. 2C**). However, other population genetic parameters, such as genetic diversity (π and θ) and divergence (D_{xy}) lack any prominent signatures (**Supplementary Fig. 19-21**).

The genome-wide patterns of genetic variation in black-bone breeds are strongly affected by introgression from commercial breeds. Hence, we focussed on the FM locus to investigate the history of the black-bone chicken breeds. As evident from the elevated F_{ST}, the major axis of genetic variation in Dup1 and Dup2 regions separates the BBC from non-black-bone chicken (**Supplementary Fig. 22**). Notably, the Kadaknath chicken breed from India has very distinctive ancestry in the Flank1 region not shared with other black-bone breeds (**Supplementary Fig. 23-24)**. We screened the genome for selection signatures to identify all such Kadaknath-specific regions by comparing Kadaknath with Chinese BBC. Chinese black-bone breeds have diverged to differing extents from Kadaknath and have limited sample sizes. Hence, we combined the individuals from XBBC, LCEM, and LCMY breeds into a single population representative of Chinese black-bone (CHIN) chicken and JETB, PENC, and GOLD into another population representative of Kadaknath (KADK).

Genome-wide signatures of selection in Kadaknath chicken

Despite the sizeable geographic separation, the genome-wide mean F_{ST} between KADK and CHIN is 0.16, making it amenable to identifying selection signatures (see **Fig. 3A** and **Table 4**). Pairwise comparison between KADK and CHIN revealed a total of 137 genic regions in the top 1% F_{ST} windows **(**see **Table 5** and **Supplementary Fig. 25-57).** We further shortlisted candidates using the additional criteria that genetic diversity (π and θ) and iHS should be strikingly different between KADK and CHIN. Two prominent signatures of selective sweeps in the KADK breed are evident on the 20th chromosome in the vicinity of the Dup1 region. The first region (R1) of ~70 Kb is ~0.7 Mb away from Dup1, and the second region (R2) of ~300 Kb is ~160 Kb away from Dup1. In both regions, the genetic diversity (π and θ) and Tajima's D (τ) are strongly reduced in KADK compared to CHIN (see **Fig. 3B** and **Supplementary Fig. 44**). The elevated iHS in KADK compared to CHIN and the positive value of XP-EHH supports a selective sweep in KADK. This region also occurs in the top 10% genome-wide D_{xy} windows (see **Supplementary Fig. 58**).

The first sweep region (R1) contains three genes with non-synonymous changes: Histocompatibility Minor 13 (*HM13*), Bactericidal/permeability-increasing protein-like (*BPIL*), and Targeting protein for Xklp2 (*TPX2*). In the second region (R2), Deoxynucleotidyltransferase Terminal Interacting Protein 1 (*DNTTIP1*), Acyl-CoA Thioesterase 8 (*ACOT8*), and solute carrier family 12 member 5 (*SLC12A5*) genes contain non-synonymous changes. The TELO2-interacting protein one homolog (*TTI1*), which occurs

near the second sweep region (R2), also has a non-synonymous change (**Fig. 3C** and **Table 6**). The most striking differences were in the *BPIL* gene, which has three fixed sites with non-synonymous changes. Of these changes, two alleles are unique to the KADK population (see **Supplementary Fig. 59**). The first non-synonymous change (C**G**G->C**A**G, R->Q) occurs at position 69 in exon-2. The second non-synonymous change (**G**AG->**A**AG, E->K) is within the first BPI superfamily protein domain at position 159 in exon-4. The third and last non-synonymous change (C**G**C->C**A**C, R->H) is in the second domain at position 405 in exon-11. We also found that *HM13, SLC12A5,* and *DNTTIP1* have novel Kadaknath-specific changes.

Similar to the sweep in KADK, we found two genomic regions with signatures of a sweep in the CHIN population (see **Table 7**). The sweep region on Chr 4 contains the *PCDH7* gene, and the region on Chr 9 contains the *COL4A3* and *MFF* genes (see **Supplementary Fig. 28, 33**). However, none of these genes have any non-synonymous fixed differences.

Kadaknath morphs are like peas in a pod

The three morphs (JETB, PENC, and GOLD) of Kadaknath are genetically very similar and lack morph-specific population structure. However, we found a strong population structure between the samples from Bhopal and Jhabua (see **Supplementary Fig. 6, 7, and 60-62**). Three of the chickens sampled from Bhopal appear as a distinct cluster (BH1), and the other three from Bhopal (BH2) seem to share ancestry with Jhabua (JHA) and BH1 (see **Supplementary Fig. 7**). We found the mean genome-wide differentiation between Jhabua and Bhopal (0.11 to 0.13) to be higher than between BH1 and BH2 (0.07) (see **Supplementary Fig. 63 and Table 4**). The mean pairwise genome-wide differentiation between morphs (F_{ST}) ranged from 0.06 to 0.09 and was less than that between Jhabua and Bhopal (see **Fig. 4**). Hence, our sampling of KADK morphs from both Jhabua and Bhopal can overcome the confounding population stratification.

Comparing KADK morphs found that PENC vs. JETB has 1,235 fixed differences, GOLD vs. JETB has 691 differences, and GOLD vs. PENC has 242 differences (see **Table 8**). In the entire genome, we found a single allele private to PENC; no such private alleles occur in the JETB and GOLD morphs. The PENC private allele occurs on Chr 11 in a region ~57Kb upstream of the *CDH5* (Cadherin 5) gene in a region of elevated evolutionary conservation indicative of a putative cis-regulatory locus (see **Supplementary Fig. 64**). We found only one non-synonymous fixed difference between GOLD and PENC in the *NAE1* gene within a

high F_{ST} region on Chr 11. In the comparison between PENC and JETB, we found one non-synonymous fixed difference each in the genes *USF3* (Chr 1), *SYCE3* (Chr1), *AGT* (Chr 3), and *IPCEF1* (Chr 3) (see **Table 8**). However, none of these alleles are unique to the KADK breed.

Introgression between black-bone and commercial breeds

Interbreeding black-bone chicken with commercial breeds to increase productivity is a common practice. Hence, we investigated genome-wide patterns of gene flow between commercial breeds (WHLL, RIRL, CWLH, and BROC) and black-bone chicken by comparing CHIN and KADK using ABBA-BABA tests. Our data suggest gene flow between KADK and all four commercial breeds (**Fig. 5A** and **Supplementary Fig**. **65**). However, we find limited gene flow between CHIN and leghorn breeds (WHLL and CWLH) compared to CHIN and two other commercial breeds (BROC & RIRL) (**Fig. 5A)**. We subsequently identified genomic regions strongly affected by gene flow based on non-overlapping-window estimates of Patterson's D-statistics and pairwise estimates of $F_{ST.}$ We found fewer genomic windows with an excess of ABBA sites supporting gene flow between CHIN and commercial breeds other than BROC (CWLH ~300Kb, WHLL ~400Kb, RIRL ~550Kb) compared to windows with an excess of BABA sites supporting gene flow between KADK and commercial breeds other than BROC (CWLH ~1.8Mb, WHLL ~1.75Mb, RIRL ~1.25Mb). However, gene flow between BROC and CHIN occurs in ~500Kb, and between BROC and KADK occurs in ~300Kb (see **Table 9)**. Several regions with an excess of ABBA or BABA sites are common for different commercial breeds (see **Table 10**).

The most conspicuous example of gene flow between KADK and commercial breeds (except BROC) occurs on Chr 4 in a ~100Kb region with a sharp D-statistic trough caused by a large number of BABA sites (see **Fig. 5B, 5C, 5D,** and **Supplementary Fig. 66-68**). The entire gamma-aminobutyric acid type A receptor alpha2 subunit (*GABRA2*) gene occurs within this region and contains three synonymous fixed differences between KADK and CHIN. These differences also occur between CHIN and commercial breeds (CWLH, WHLL, and RIRL) (see **Supplementary Table 11**). Notably, deletions of 3, 3, and 11 bases occur in intron 2, 4, and 4 respectively of the *GABRA2* gene at a higher frequency in KADK than CHIN BBC (see **Supplementary Fig. 69-72**). Similar to KADK, the commercial breeds CWLH, WHLL, and RIRL contain these deletions, which occur at low frequency in BROC (see **Supplementary Fig. 73-76**). Another region of ~50Kb on Chr 4 provides a contrasting example of gene flow

between KADK and two commercial breeds (WHLL and RIRL) but not two other commercial breeds (CWLH and BROC) (see **Supplementary Fig. 77-80**). This region contains the gene *C4orf33*, which has six fixed differences between CHIN and KADK (see **Table 11**). The KADK allele occurs in WHLL and RIRL, while the CHIN allele occurs in BROC and CWLH. We found another ~100Kb genomic region with an excess of BABA sites on Chr 11 as a candidate for gene flow between KADK and commercial breeds (except BROC) (see **Supplementary Fig. 81-84**). This region contains the *PRMT7* gene, which has two synonymous and one non-synonymous fixed difference between KADK and CHIN (see **Table 11**).

Discussion

All BBC breeds share a common origin

As part of this study, we have generated the first whole-genome dataset of the Kadaknath breed spanning all three morphs. Hence, by comparing our data with public datasets, we can evaluate the relationship of all BBC breeds. Our comparison of the FM locus junction region in genome sequencing data from BBC breeds conclusively establishes a common origin for the complex chromosomal rearrangement[34]. Although each BBC breed has considerable genetic distinctiveness, population genetic analysis supports the common heritage of all BBC breeds and reveals a trend of isolation by distance (IBD). While IBD patterns are well established in Chinese native chicken breeds[35], our dataset spans BBC breeds from India, China, Indonesia, and Korea. The sample size of some BBC breeds in our analysis is limited, and the sampling does not cover the entire geographic distribution of some breeds. However, our analysis achieves genome-wide pan-BBC breed sampling by including all the major breeds from Asia.

Within the Kadaknath breed, all three morphs are genetically similar and lack a population sub-structure. The Kadaknath breed is genetically closest to the BBC chicken from Tibet and suggests that the old Tibet-Nepal salt trade route may have facilitated the spread of BBC. Interestingly, the Tibetan BBC is genetically more similar to Kadaknath than some Chinese BBC breeds. Unfortunately, we lack BBC samples from Nepal. However, we believe BBC poultry in Nepal is Kadaknath chicken recently imported from India. Hence, the ancestral stock of the BBC spread across Asia may not be currently available. Changes in trade routes and the introduction of commercial poultry breeds limit our ability to trace the historical prevalence of BBC breeds.

Did black-bone chicken originate in India?

The lack of data from India and the considerable interest in BBC breeds in Europe, China, and Korea has meant that studies have focussed mainly on non-Indian BBC. Hence, Southern China and Tibet are considered the source of all BBC breeds[36,37]. However, despite a traditionally restricted geographic distribution, the Kadaknath breed has nucleotide diversity comparable to Chinese BBC breeds and much higher than the Korean BBC. The number of private alleles identified in Kadakanth is also greater than in Chinese and Korean BBC breeds. Moreover, the mean genome-wide F_{ST} of 0.11 between Jhabua and Bhopal is comparable to the differentiation between some Chinese BBC breeds. The presence of three distinct morphs within Kadaknath chicken suggests phenotypic diversity derived either from commercial breeds or existing variations within Kadaknath. The extent of phenotypic diversity within BBC breeds from China and Korea is not documented to allow a fair comparison. The high genetic diversity in Kadaknath supports the potential origin of all BBC in Jhabua, India. The geographic proximity of Jhabua to the sympatric distribution of the Red junglefowl and the Grey junglefowl suggests that BBC was derived soon after the domestication of chicken[1,3,38]. The export of black chicken from India to Africa in ~1600 AD also supports that BBC was present in ancient India[2].

While we cannot conclude whether BBC had an Indian, Chinese, or more Southeast Asian origin, our data suggest that any of these sources are plausible. More widespread geographic sampling and analysis of allele sharing patterns may provide a definitive answer regarding the origin of BBC. Unlike the initial domestication of the chicken, which may have occurred independently[39,40] in several locations, the BBC has a single origin linked to the rearrangement at the FM locus. Irrespective of the source of all BBC, we identify several Kadaknath-specific genetic changes caused due to selection and introgression events. Hence, the genetic distinctiveness of the Kadaknath breed has long diverged from other BBC breeds and is a result of its unique heritage sustained by the Bhil and Bhilala tribal communities of Madhya Pradesh.

Selective sweep near FM locus may be a consequence of linkage

Our analysis identified two genomic regions (R1 and R2) in the vicinity of the FM locus with prominent signatures of a selective sweep in the KADK population. The most promising candidate gene in this region that may have been the focus of selection is the *BPIL* gene which has accumulated three non-synonymous changes. The *BPIL* gene is part of the innate immune defense system and binds to and neutralizes LPS from the outer membrane of Gram-

negative bacteria[41–43]. In the chicken genome, five *BPI*-like genes occur near *BPI*[44], which works as a bactericidal protein and helps arrest bacterial growth and phagocytosis by neutrophils[45,46]. In chickens, defense against gram-negative bacteria may be especially important as they lack the complement *C9* gene required to effectively eliminate pathogens through the Membrane Attack Complex (MAC) formation[47]. Notably, another BBC breed from Korea, YOSK, has signatures of selection at the TLR4 gene involved in detecting gram-negative bacteria[20]. Hence, various immune genes may have been selected in different BBC breeds to protect against gram-negative bacteria.

Several other genes with non-synonymous changes occur in R1 and R2 genomic regions. Although we cannot identify the functional consequences of these changes, the strong signature of selection and KADK-specific protein-coding alterations suggest phenotype-altering ability. The proximity of R1 and R2 regions to the FM locus indicates that selective sweep may be due to close physical linkage with the FM locus. Such co-selection of traits due to linkage and pleiotropy has occurred during animal domestication, including in chicken[48,49]. Hitchhiking of alleles can increase the frequency of both beneficial and mildly deleterious alleles. Whether the changes in KADK are beneficial or deleterious will need to be investigated in future studies.

The genomic co-occurrence of agronomically important traits in domesticated plants is also known[50–53]. The alleles causing phenotypic changes can occur in closely linked genes and will undergo selection for any of the phenotypes being favored by the breeding process. Notably, the selective sweep is found only in KADK and not in other BBC breeds. Hence, the genetic variants and the selective sweep may represent recent events after the other BBC breeds separated from KADK. Comparing different BBC breeds provides snapshots of the selection process and the build-up of genetically linked co-selected allelic changes and may help understand the hitchhiking process during domestication.

Introgression between commercial and black-bone chicken breeds

Introgression has been an important process during chicken domestication and is relevant in determining the genetic constituency of present-day breeds[54–57]. Indigenous breeds such as native chicken, including BBC, have been strongly affected by introgression from commercial breeds[35]. We find that the Kadaknath breed is no exception and has experienced gene flow with commercial breeds. Interestingly, while we find evidence of gene flow between Kadaknath and commercial breeds such as layer (WHLL and CWLH) and dual-

purpose (RIRL) breeds, BBC breeds from China seem to have had limited gene flow with these breeds. However, the magnitude of gene flow between Kadaknath and BROC seems comparable to the gene flow between the Chinese BBC and BROC. The difference in the patterns of introgression between Kadaknath and Chinese BBC breeds may reflect country-specific poultry practices. For instance, the small population of Kadaknath[58] may have spurred the increased use of layer breeds to expand the population. However, other BBC breeds with large populations may not require the extensive use of layer breeds.

The cross-breeding of commercial breeds with BBC to increase productivity has resulted in introducing commercial breed-specific alleles into BBC breeds. While this disrupts the genetic constituency of the BBC breeds, it also provides a novel opportunity for trait mapping by identifying the genomic regions that have introgressed. For instance, the feather pecking behavior of layers also occurs in Kadaknath and may result from introgression from commercial breeds. Interestingly, we identify gene flow between Kadaknath and commercial breeds at the *GABRA2* (Gamma-aminobutyric acid type A receptor alpha2 subunit) locus, whose expression downregulation is associated with feather pecking in birds[59]. The intronic deletions that we identified in the *GABRA2* locus, which occur in Kadaknath and commercial breeds, may be involved in regulating this gene. Another prominent signature of introgression occurs at the *PRMT7* gene, which maintains muscle stem cell regenerative capacity[60], and its deficiency causes impairment of social behaviors[61]. The candidate regions we identify are good targets for genetic manipulations in chickens to evaluate their phenotype-altering abilities. Our approach of using Kadaknath to identify genomic regions of introgression from commercial breeds for trait mapping may broadly apply to other native chicken breeds.

Genome-wide data generated as part of this study and comparative analysis with other BBC breeds establishes the genetic uniqueness of Kadaknath. We also identify specific genes with signatures of selection that are likely to be responsible for the Kadaknath-specific phenotypes. The gene flow between Kadaknath and various commercial chicken breeds has identifiable signatures in the genome and quantifies the magnitude of genetic erosion. Hence, our results have important implications for understanding the genetic basis of Kadaknath-specific phenotypes and their conservation. Co-selection of genes that are in linkage, as shown in Kadaknath, is widespread in domestic species. Such clustering of linked alleles under selection is favored in low recombination regions and near chromosomal rearrangements. Our work exemplifies the interaction of artificial selection and chromosomal

rearrangement linked traits in domesticated species. Kadaknath is also a well-known example of the role played by tribal communities in domestication and conservation.

Materials and Methods

Population sampling

We purchased the meat of six individuals (two individuals each from Jetblack, Pencil, and Golden morphs (**Supplementary Fig. 1**)) of Kadaknath from an FSSAI (Food Safety and Standards Authority of India) licensed shop in Bhopal, Madhya Pradesh, India. We procured two other individuals with a black-bone phenotype from the same FSSAI licensed shop to examine hybrids. One had a Golden-Pencil external appearance, and another was completely white on the exterior. We also sampled one non-black-bone chicken from the same shop to determine the genetic relationship of Kadaknath with native-village chicken reared in the same poultry (see **Supplementary Fig. 1**). To reduce the chances of inter-breeding and have relatively pure Kadaknath samples, we obtained three (one individual each from Jetblack, Pencil, and Golden morphs) additional Kadaknath breed chickens from Jhabua, Madhya Pradesh. Whole-genomic DNA with high purity and quality was extracted from the tissue samples using DNeasy Blood & Tissue Kits (QIAGEN). We generated >25x coverage whole-genome short-read paired-end data (using Illumina Novaseq) for all 12 individuals sampled (see **Table 1**).

To evaluate the relationship of Kadaknath chicken globally, we compared our dataset with publicly available chicken resequencing data. High-quality resequencing datasets of 89 chicken individuals from other BBC, commercial chicken lines, and other distinctive chicken breeds with a coverage >20x were selected and obtained from ENA (Europen Nucleotide Archive) and (KNABIC) Korean National Agricultural Biotechnology Information Center (https://nabic.rda.go.kr/). Hence, we analyze a dataset of 101 individuals (89 public + 12 sequenced as part of this study) from different breeds (see **Table 2** for more details)[20,21,25,40,62–67].

Read mapping and variant calling

We mapped the paired-end raw reads of 101 individuals to the chicken genome assembly (Gallus_gallus.GRCg6a) using the BWA[68] (Burrows-Wheeler Aligner) v0.7.17-r1188 read mapper. We added the read group information using Picard tools and removed duplicate reads in all 101 individual bam files (https://github.com/broadinstitute/picard). The Single Nucleotide Polymorphisms (SNPs) identified by both variant callers (bcftools[69],[70] v1.9 and

FreeBayes[71] v1.0.0) were used for subsequent analysis (see **Supplementary File S1** for details of settings used).

FM-locus junction identification

The FM locus consists of a complex chromosomal rearrangement composed of two duplicated regions (Dup1 and Dup2) separated by an intermediate (Int) region. To identify the base-pair level positions of Dup1, Dup2, and Int regions, we compared the short-read-coverage of black and non-black chicken in 1Kb windows along chromosome 20. We shortlisted adjacent windows with drastically different read coverage in black individuals but not in non-black individuals. The base-pair level coordinates of Dup1 and Dup2 in the Galgal6 genome were narrowed down further using coverage estimates in 1bp windows. Dup1 starts at 20:10766727-10766827 and ends at 20:10894100-10894200. Dup2 occurs further along the chromosome and starts at 20:11306637-11306737 and ends at 20:11477385-11477551. The read coverage at these junctions consistently differed between all black-bone and non-black-bone chicken. In addition to the Dup1, Dup2, and Int regions, we defined ~500 Kb flanking regions as Flank1 (20:10263555-10766771) and Flank2 (20:11477502-11980000).

Black-bone-specific FM-locus junction

The junction between the rearranged regions in the BBC does not occur in the Galgal6 genome (**Supplementary Fig. 85-86**). Hence, we searched for the FM-locus junction in the Korean BBC PacBio data from ENA (SRR6189090). Based on our search of PacBio reads that mapped to the Gallus_gallus.GRCg6a genome, we shortlisted reads that simultaneously aligned to two of the five (Flank1, Dup1, Int, Dup2, and Flank2) genomic regions we have defined. We found several reads spanning Flank1-Dup1, Dup1-Int, Int-Dup2, and Dup2-Flank2, which is unsurprising because these regions are adjacent.

We also found five reads (SRR6189090.111279, SRR6189090.56386, SRR6189090.880702, and SRR6189090.387043 and SRR6189090.54824) that span across both Dup1 and Dup2. Dup1 and Dup2 are far apart in their genomic location in Galgal6 assembly. Hence, these reads support a rearrangement that leads to junctions between these two regions. Of these five, three reads (SRR6189090.111279, SRR6189090.880702, and SRR6189090.387043) support the junction START-DUP1-END-END-DUP2-START, and the other two reads (SRR6189090.54824 and SRR6189090.56386) support the junction END-DUP1-START-

START-DUP2-END. The read coverage at these junctions consistently differed between all black-bone and non-black-bone chicken. We identified 34 out of 101 samples analyzed as BBC based on the read coverage across the FM-locus junctions.

Principal component analysis, admixture, and phylogeny

Principal components analysis (PCA) was performed using PCAngsd[72] based on genotype likelihood estimates from ANGSD[73] (Analysis of Next Generation Sequencing Data) v0.935 (see **Supplementary File S1** for details). PCA was done both genome-wide and FM locus region-wise (Flank1, Dup1, Int, Dup2, and Flank2) for the 34 black-bone and all 101 chicken individuals**.** We investigated (Clumpak[74] identified seven as the best K) genome-wide admixture using NGSadmix[75]. We constructed SNP-based maximum-likelihood trees with 1000 bootstraps using SNPhylo[76] v20180901 on genome-wide data except for Z, W, Mt, and unplaced scaffolds. The Mt genome haplotype median-joining networks were constructed using PopART[77] and SplitsTree[78].

Population genetic analysis

We calculated population genetic statistics genome-wide to identify signatures of selection. However, to avoid false positives, we excluded genomic regions (50 Kb windows with <80 percent callable sites) with poor callability. We used the CallableLoci walker of GATK[79] on the bam files to quantify callability. Furthermore, we evaluated the density of repeat elements in genomic regions with signatures of selection to rule out the possibility of false positives (see **Supplementary Fig. 87-89**). We used stringent coverage and quality criteria (-fold 1, -dosaf 1, -baq 1, -C 50, -minInd 3, 6 or 9, -minMapQ 30, -minQ 20) in ANGSD to calculate all population genetic statistics. The list of individuals in each population and the population pairs compared is in **Table 12**. We found that at least 20,000 of the 21,659 50Kb windows had sufficient high-quality data in all populations.

Within and between population statistics

We used the folded site frequency spectrum (SFS) approach implemented in ANGSD to calculate genome-wide population-specific estimates of π (the average pairwise differences), Watterson's θ (the average number of segregating sites), τ (Tajima's D), and Fu and Li's D. We estimated inter-population genomic differentiation (F_{ST}) and divergence (D_{xy}) using ANGSD and popgenWindows python script

(https://github.com/simonhmartin/genomics_general) for each population pair. The F_{ST} estimates from the two methods were highly correlated (ρ= 0.98, p-value < 2.2 e-16)(see **Supplementary Fig. 90-91**). We defined F_{ST} outlier regions as 50Kb windows in the top 1 percent of the genome-wide estimates and merged adjoining windows using bedtools. Similarly, the 50Kb windows in the top 10 percent of the genome-wide estimates of D_{xy} were deemed to have high levels of divergence. We identified fixed sites as those sites with F_{ST} > 0.9. The effect of each variant was determined using snpEff[80]. We estimate haplotype-based statistics iHS and XP-EHH in the rehh[81] R package with option polarised=FALSE using genotype data phased with SHAPEIT[82].

Introgression and Patterson's D-statistic

We calculated the number of ABBA and BABA sites and the D-statistic using ANGSD (genome-wide) and in 50Kb windows with ABBABABAwindow python script (https://github.com/simonhmartin/genomics_general). We use the jackKnife.R script to estimate the Z score. We identified genomic regions with introgression as 50Kb windows with at least 25 ABBA or BABA sites and between population F_{ST} in the top 5 percentile in two population pairs and the bottom 95 percentile in the other population pair.

Acknowledgment

We thank the Council of Scientific and Industrial Research for fellowship to SSS and the University Grants Commission for supporting AS with a Ph.D. scholarship. The Department of Biotechnology, Ministry of Science and Technology, India (Grant no. BT/11/IYBA/2018/03) and Science and Engineering Research Board (Grant no. ECR/2017/001430) provided funds used to generate primary sequencing data published in this article and computational resources (i.e., Har Gobind Khorana Computational Biology cluster) used.

Author contributions

SSS and AS wrote the manuscript with major inputs from NV. SSS analyzed the data along with AS. SSS also generated all the new sequencing data presented in this study. All authors reviewed the manuscript.

Competing interest statement

None to declare

Availability of data

This study's primary sequencing data are available from the SRA under BioProject# PRJEB51457. All **Supplementary information**, including other associated data and scripts

used for analysis, are provided in an easy-to-browse format: **https://github.com/ceglabsagarshinde/Kadaknath_Project**.

References

1. Wang, M. S. *et al.* 863 genomes reveal the origin and domestication of chicken. *Cell Res*. **30**, 693 (2020).

2. Tixier-Boichard, M., Bed'Hom, B. & Rognon, X. Chicken domestication: from archeology to genomics. *C. R. Biol.* **334**, 197–204 (2011).

3. Lawal, R. A. & Hanotte, O. Domestic chicken diversity: Origin, distribution, and adaptation. *Anim. Genet.* **52**, 385–394 (2021).

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

5. Cogburn, L. A. *et al.* Functional genomics of the chicken--a model organism. *Poult. Sci.* **86**, 2059–2094 (2007).

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

7. Kralik, G., Kralik, Z., Grčević, M. & Hanžek, D. Quality of Chicken Meat. *Anim. Husb. Nutr* 63–94 (2016) doi:10.5772/intechopen.72865.

8. Mir, N. A., Rafiq, A., Kumar, F., Singh, V. & Shukla, V. Determinants of broiler chicken meat quality and factors affecting them: a review. *J. Food Sci. Technol.* **54**, 2997–3009 (2017).

9. Jaturasitha, S., Srikanchai, T., Kreuzer, M. & Wicke, M. Differences in Carcass and Meat Characteristics Between Chicken Indigenous to Northern Thailand (Black-Boned and Thai Native) and Imported Extensive Breeds (Bresse and Rhode Island Red). *Poult. Sci.* **87**, 160–169 (2008).

10. Muroya, S., Tanabe, R. I., Nakajima, I. & Chikuni, K. Molecular characteristics and site specific distribution of the pigment of the silky fowl. *J. Vet. Med. Sci.* **62**, 391–395 (2000).

11. Bateson, B. W., -a, F., Punnett, R. C. & Contents, M. The inheritance of the peculiar pigmentation of the silky fowl. *J. Genet. 1911 13* **1**, 185–203 (1911).

12. Dorshorst, B. *et al.* A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. *PLoS Genet.* (2011).

13. Dharmayanthi, A. B. *et al.* The origin and evolution of fibromelanosis in domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds. *PLoS One* (2017).

14. Shinomiya, A. *et al.* Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (fibromelanosis) in silky chickens. *Genetics* (2012).

15. Tu, Y. gang, Sun, Y. zhen, Tian, Y. gang, Xie, M. yong & Chen, J. Physicochemical characterisation and antioxidant activity of melanin from the muscles of Taihe Black-bone silky fowl (Gallus gallus domesticus Brisson). *Food Chem.* **114**, 1345–1350 (2009).

16. Dou, T. *et al.* Integrative analysis of transcriptomics and metabolomics to reveal the melanogenesis pathway of muscle and related meat characters in Wuliangshan black-boned chickens. *BMC Genomics* **23**, 1–15 (2022).

17. Tian, Y. *et al.* Determination of carnosine in Black-Bone Silky Fowl (Gallus gallus domesticus Brisson) and common chicken by HPLC. *Eur. Food Res. Technol.* **226**, 311–314 (2007).

18. Sharma, R. *et al.* An attempt to valorize the only black meat chicken breed of India by delineating superior functional attributes of its meat. *Sci. Reports 2022 121* **12**, 1–12 (2022).

19. Tian, Y. *et al.* Composition of fatty acids in the muscle of black-bone silky chicken (Gallus gellus demesticus brissen) and its bioactivity in mice. *Food Chem.* **126**, 479–483 (2011).

20. Cho, Y., Kim, J. Y. & Kim, N. Comparative genomics and selection analysis of Yeonsan Ogye black chicken with whole-genome sequencing. *Genomics* **114**, (2022).

21. Sohn, J. Il *et al.* Whole genome and transcriptome maps of the entirely black native Korean chicken breed Yeonsan Ogye. *Gigascience* (2018).

22. Buranawit, K., Chailungka, C., Wongsunsri, C. & Laenoi, W. Phenotypic characterization of Thai native black-bone chickens indigenous to northern Thailand. *Thai J. Vet. Med.* **46**, 547–554 (2016).

23. Zhu, W. Q. *et al.* Molecular genetic diversity and maternal origin of Chinese black-bone chicken breeds. **13**, 3275–3282 (2014).

24. Kriangwanich, W. *et al.* Consistency of dark skeletal muscles in Thai native black-bone chickens (Gallus gallus domesticus). *PeerJ* **9**, (2021).

25. Li, D. *et al.* Breeding history and candidate genes responsible for black skin of Xichuan black-bone chicken. *BMC Genomics* **21**, (2020).

26. Huang, X. *et al.* Genome-wide genetic structure and selection signatures for color in 10 traditional Chinese yellow-feathered chicken breeds. *BMC Genomics* **21**, (2020).

27. Haunshi, S. & Prince, L. L. L. Kadaknath: a popular native chicken breed of India with unique black colour characteristics. *https://doi.org/10.1080/00439339.2021.1897918* **77**, 427–440 (2021).

28. Sehrawat, R. *et al.* First report on better functional property of black chicken meat from India. *Indian J. Anim. Res.* **55**, 727–733 (2021).

29. Haunshi, S. *et al.* Characterization of two Indian native chicken breeds for production, egg and semen quality, and welfare traits. *Poult. Sci.* **90**, 314–320 (2011).

30. Sahu Senior Scientist, N. *et al.* Woman empowerment and uplifting of tribal farmers through Kadaknath poultry farming in Dantewada district of Chhattisgarh. *~ 315 ~ J. Entomol. Zool. Stud.* **7**, 315–319 (2019).

31. Thakur, M. S. *et al.* Evaluation of Kadaknath Chicken for Coccidial Resistance by Oocyst Count, Lesion Scoring and Oocyst Index in Eimeria Tenella Infection . *J. Anim. Res.* **5**, 579 (2015).

32. Jena, P. *et al.* Reproductive Performance and Egg Quality Traits of Kadaknath in

Intensive Managemental Condition under Hot and Humid Climate. *Int. J. Livest. Res.* 1 (2018) doi:10.5455/IJLR.20180131054034.

33. Shanmathy, M. *et al.* Comparative Assessment on Performance of Aseel and Kadaknath in Hot and Humid Conditions in Tropics. *Int. J. Curr. Microbiol. Appl. Sci.* **7**, 2156–2165 (2018).

34. Dharmayanthi, A. B. *et al.* The origin and evolution of fibromelanosis in domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds. *PLoS One* **12**, (2017).

35. Nie, C. *et al.* Genome-Wide Single-Nucleotide Polymorphism Data Unveil Admixture of Chinese Indigenous Chicken Breeds with Commercial Breeds. *Genome Biol. Evol.* **11**, 1847–1856 (2019).

36. Zhu, W. Q. *et al.* Molecular genetic diversity and maternal origin of Chinese black-bone chicken breeds. *Genet. Mol. Res.* **13**, 3275–3282 (2014).

37. Zhang, T., Du, W., Lu, H. & Wang, L. Genetic Diversity of Mitochondrial DNA of Chinese Black-bone Chicken. *Brazilian J. Poult. Sci.* **20**, 565–572 (2018).

38. Peters, J. *et al.* The biocultural origins and dispersal of domestic chickens. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).

39. Tixier-Boichard, M., Bed'Hom, B. & Rognon, X. Chicken domestication: from archeology to genomics. *C. R. Biol.* **334**, 197–204 (2011).

40. Mariadassou, M. *et al.* Unraveling the history of the genus Gallus through whole genome sequencing. *Mol. Phylogenet. Evol.* **158**, (2021).

41. Sun, Y. Y. & Sun, L. A Teleost Bactericidal Permeability-Increasing Protein Kills Gram-Negative Bacteria, Modulates Innate Immune Response, and Enhances Resistance against Bacterial and Viral Infection. *PLoS One* **11**, (2016).

42. Holweg, A., Schnare, M. & Gessner, A. The bactericidal/permeability-increasing protein (BPI) in the innate defence of the lower airways. *Biochem. Soc. Trans.* **39**, 1045–1050 (2011).

43. Monson, M. S. *et al.* Genetic lines respond uniquely within the chicken thymic transcriptome to acute heat stress and low dose lipopolysaccharide. *Sci. Rep.* **9**, (2019).

44. Chiang, S. C. *et al.* Identification and characterisation of the BPI/LBP/PLUNC-like gene repertoire in chickens reveals the absence of a LBP gene. *Dev. Comp. Immunol.* **35**, 285–295 (2011).

45. Elsbach, P. & Weiss, J. Role of the bactericidal/permeability-increasing protein in host defence. *Curr. Opin. Immunol.* **10**, 45–49 (1998).

46. Weiss, J. Bactericidal/permeability-increasing protein (BPI) and lipopolysaccharide-binding protein (LBP): structure, function and regulation in host defence against Gram-negative bacteria. *Biochem. Soc. Trans.* **31**, 785–790 (2003).

47. Sharma, A., Gupta, S., Patil, A. B. & Vijay, N. Birth and Death in Terminal Complement Pathway. *bioRxiv* 2022.04.21.489004 (2022) doi:10.1101/2022.04.21.489004.

48. Rubin, C. J. *et al.* Whole-genome resequencing reveals loci under selection during

chicken domestication. *Nat. 2010 4647288* **464**, 587–591 (2010).

49. Webster, M. T. *et al.* Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. *BMC Genomics* **16**, 1–17 (2015).

50. Li, J. *et al.* Identification of ST1 reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. *Plant Biotechnol. J.* **20**, 1110–1121 (2022).

51. Fang, C. *et al.* Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **18**, (2017).

52. Zhang, D. *et al.* Quantitative trait loci associated with soybean tolerance to low phosphorus stress based on flower and pod abscission. *Plant Breed.* **129**, 243–249 (2010).

53. Kong, L. *et al.* Quantitative Trait Locus Mapping of Flowering Time and Maturity in Soybean Using Next-Generation Sequencing-Based Analysis. *Front. Plant Sci.* **9**, (2018).

54. Sawai, H. *et al.* The Origin and Genetic Variation of Domestic Chickens with Special Reference to Junglefowls Gallus g. gallus and G. varius. *PLoS One* **5**, e10639 (2010).

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

56. Leroy, G. *et al.* Gene diversity, agroecological structure and introgression patterns among village chicken populations across North, West and Central Africa. *BMC Genet.* **13**, 1–11 (2012).

57. Zhang, C. *et al.* Widespread introgression in Chinese indigenous chicken breeds from commercial broiler. *Evol. Appl.* **12**, 610 (2019).

58. Sharma, R. *et al.* An attempt to valorize the only black meat chicken breed of India by delineating superior functional attributes of its meat. *Sci. Reports 2022 121* **12**, 1–12 (2022).

59. Falker-Gieske, C., Bennewitz, J. & Tetens, J. The light response in chickens divergently selected for feather pecking behavior reveals mechanistic insights towards psychiatric disorders. *Mol. Biol. Rep.* **49**, 1649–1654 (2022).

60. Blanc, R. S., Vogel, G., Chen, T., Crist, C. & Richard, S. PRMT7 Preserves Satellite Cell Regenerative Capacity. *Cell Rep.* **14**, 1528–1539 (2016).

61. Lee, S. Y. *et al.* PRMT7 deficiency causes dysregulation of the HCN channels in the CA1 pyramidal cells and impairment of social behaviors. *Exp. Mol. Med.* **52**, 604–614 (2020).

62. Wang, M. S. *et al.* Genomic Analyses Reveal Potential Independent Adaptation to High Altitude in Tibetan Chickens. *Mol. Biol. Evol.* **32**, 1880–1889 (2015).

63. Ulfah, M. *et al.* Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. *BMC Genomics* **17**, (2016).

64. Roux, P. F. *et al.* Re-sequencing data for refining candidate genes and polymorphisms in QTL regions affecting adiposity in chicken. *PLoS One* **9**, (2014).

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

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

67. Dorshorst, B. *et al.* A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two duplex-comb phenotypes. *PLoS Genet.* **11**, (2015).

68. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

69. Li, H. & Barrett, J. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

70. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).

71. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. (2012).

72. Meisner, J. & Albrechtsen, A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* **210**, 719–731 (2018).

73. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

74. 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).

75. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).

76. Lee, T. H., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 162 (2014).

77. Leigh, J. W. & Bryant, D. popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

78. Huson, D. H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73 (1998).

79. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

80. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* **6**, 80–92 (2012).

81. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).

82. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods 2011 92* **9**, 179–181 (2011).

Figure Legends

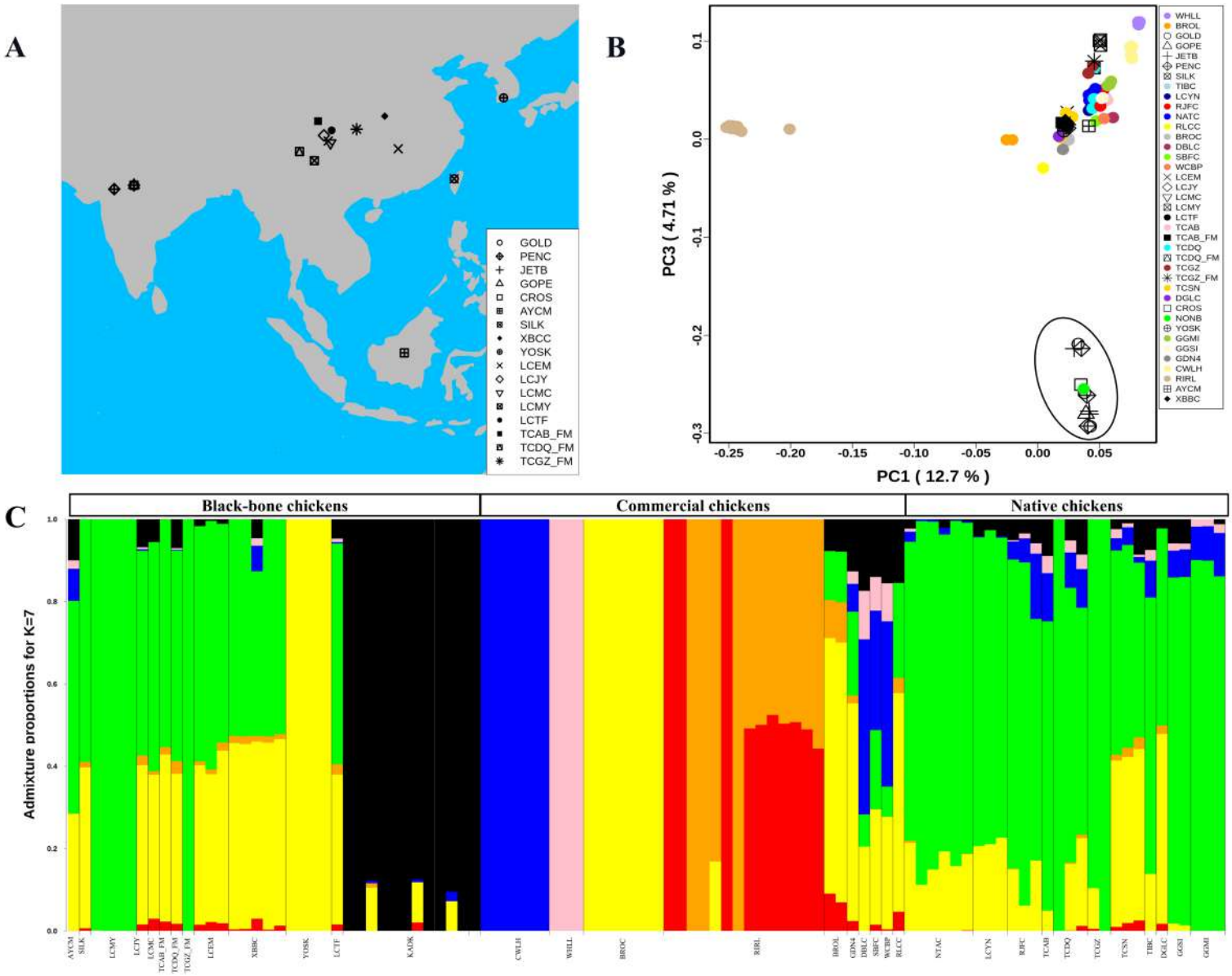
Fig. 1: Population structure analysis : (A) Geographical locations of different black-bone chicken breeds used in this study are shown on the map using different shapes. The map was generated using rworldmap, map, and mapdata R packages. **(B)** Genome-wide principal component analysis (PCA) reveals the genetic relationship of 101 chicken individuals. PC1 and PC3 explained 12.7% and 4.71% variance, respectively. Different colors show non-black-bone breeds. All black-bone breeds are shown in black with different shapes for each breed. The black oval-shaped circle represents the Kadaknath breed in a separate cluster. **(C)** Population genetic structure and individual ancestry were estimated using NGSadmix for 101 chickens from different breeds based on best K=7.

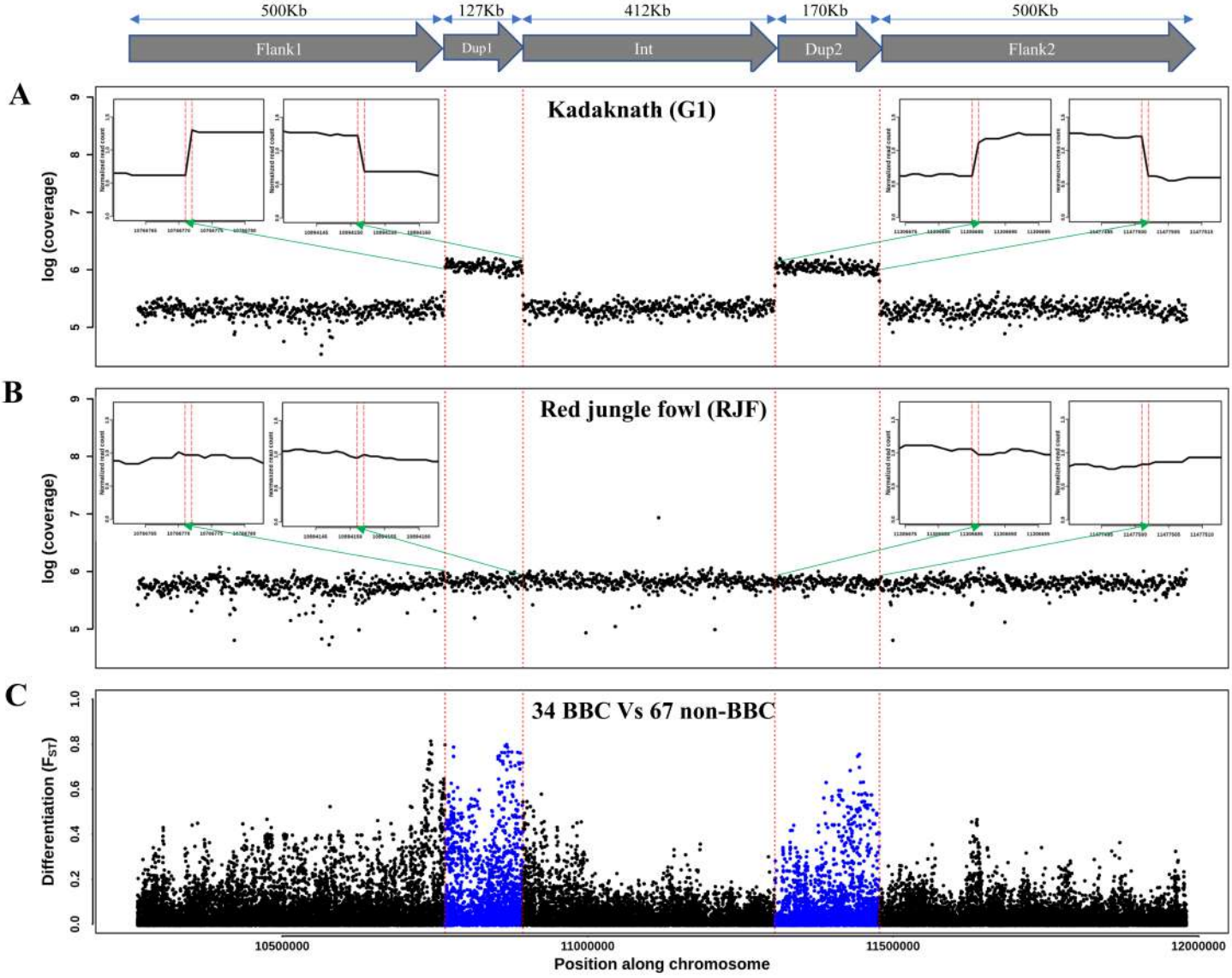
Fig. 2: Comparisons of FM locus in black-bone vs. non-black-bone chicken: Blue-colored bi-directional arrows at the top of the figure represent each region's length. Read coverage along chromosome 20 at the FM locus is shown in 1Kb sliding windows. Two duplicated genomic loci, Dup1 and Dup2, have a higher coverage in **(A)** Black-bone Kadaknath (Golden1) individuals than **(B)** non-black-bone Red junglefowl (RJF). The inset shows the normalized base-pair level read coverage at each of the junctions with 15 bp flanks. **(C**) F_{ST} landscape of FM locus between 34 black-bone chickens (BBC) and 67 non-black-bone chickens (non-BBC) is shown in a 1bp window. Dup1 and Dup2 regions are represented in blue color dots, showing high F_{ST} compared to the nearby regions.

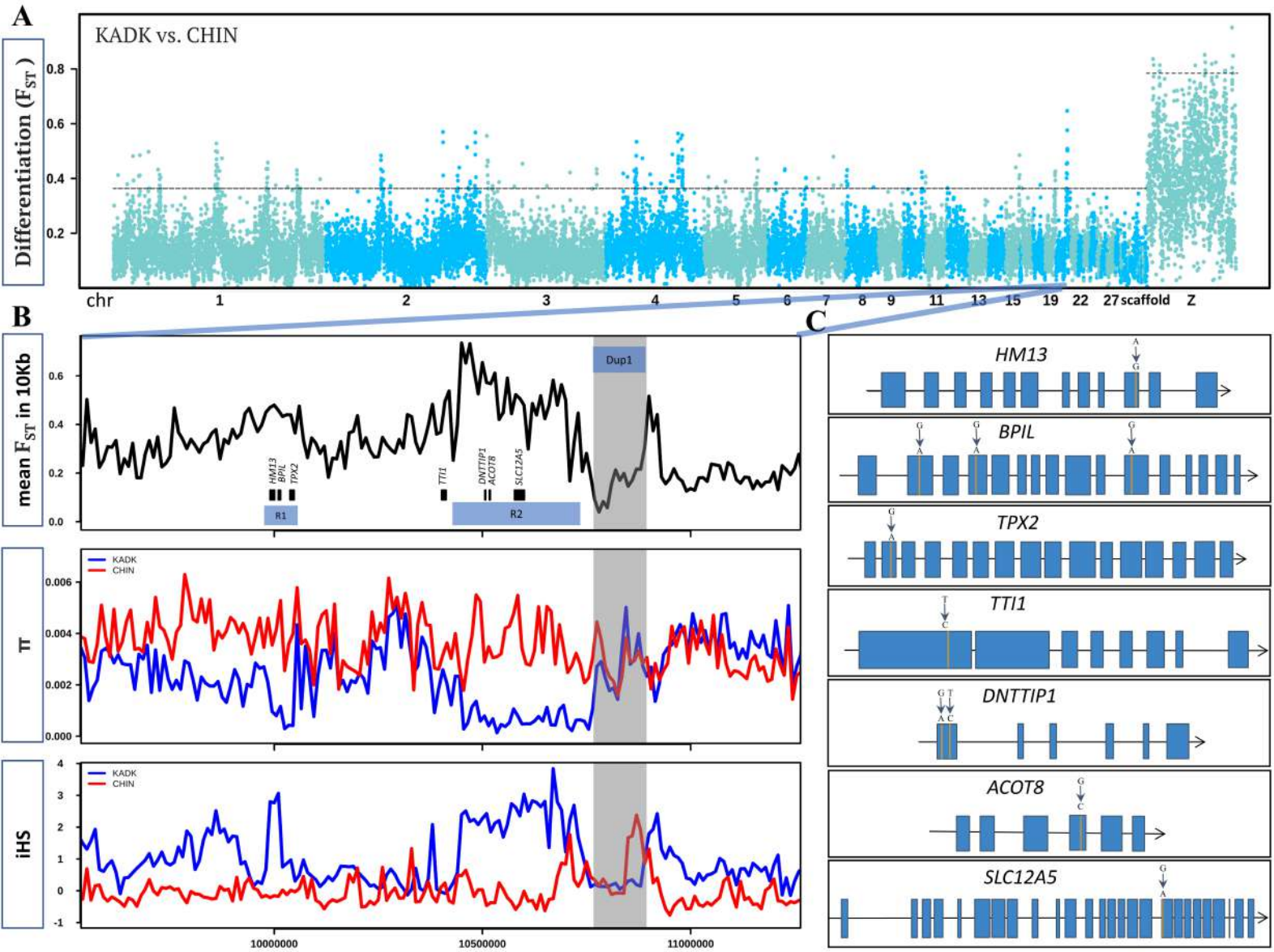
Fig. 3: (A) Genome-wide landscape of pairwise genetic differentiation (F_{ST}) between KADK and CHIN in 50Kb non-overlapping windows. The dark slate gray and deep sky blue colors represent the alternative chromosomes. The dotted horizontal black line marks the 99th percentile outlier of estimated F_{ST} for autosome and the Z chromosome, respectively. **(B)** The highest F_{ST} region with a prominent pattern occurs on chromosome 20. The panels from top to bottom show estimates of F_{ST} in 10Kb non-overlapping windows within population pairwise genetic diversity (π) and integrated haplotype score (iHS) with major alleles as ancestral and minor allele as derived. The genes are represented as black boxes with names. The transparent grey color highlights the Dup1 region in all panes in the B panel, while R1 and R2 represent the selective sweep regions. **(C)** Non-synonymous changes (shown above the exon) within genes in the regions R1 and R2 are denoted by vertical orange lines. Blue boxes are the exons connected by black lines representing introns.

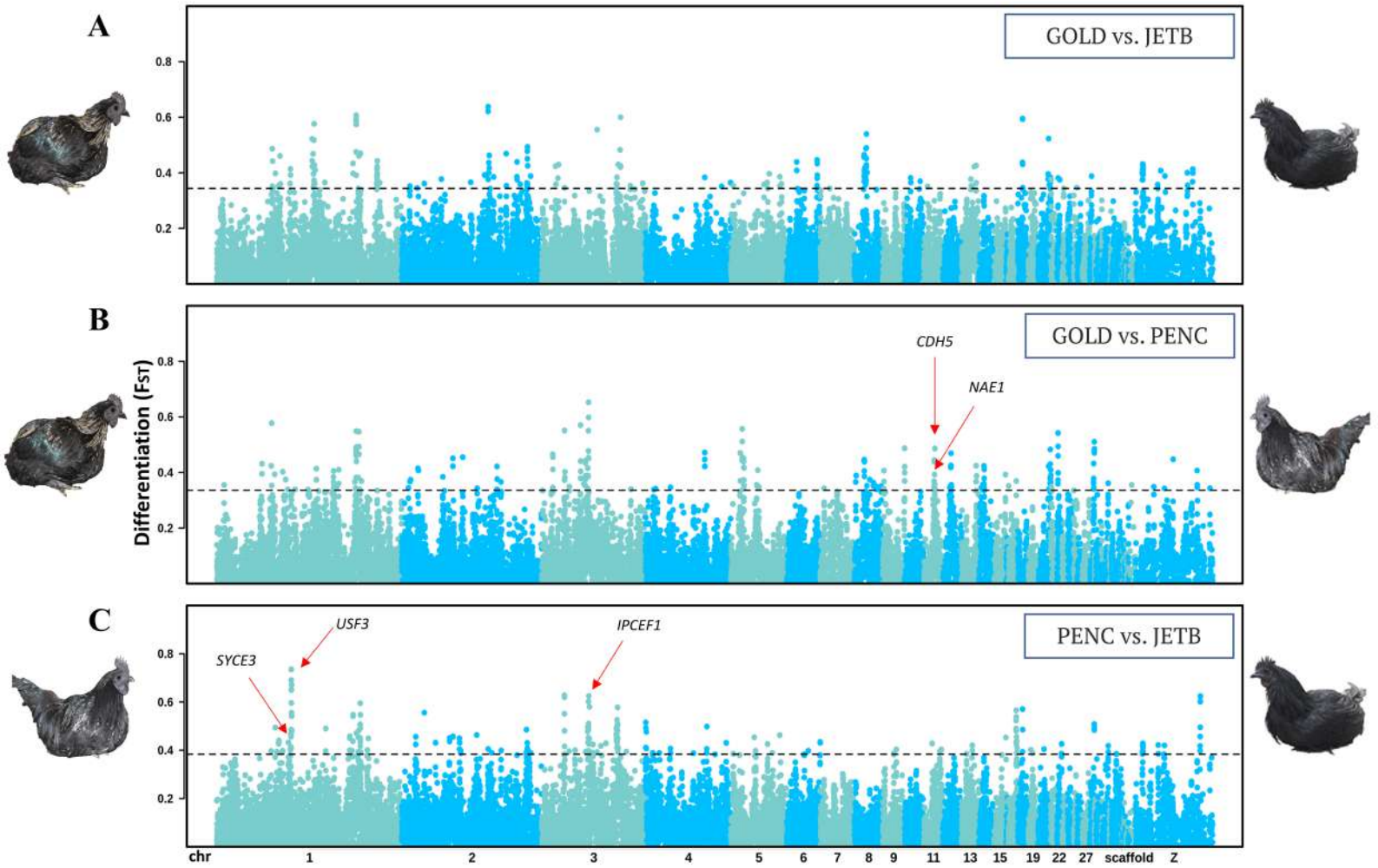
Fig. 4: Genomic landscape of population genetic differentiation (F_{ST}) between Kadaknath morphs: Pairwise genetic differentiation (F_{ST}) in 50Kb non-overlapping windows across the genome. **(A)** Golden (GOLD) vs. Jetblack (JETB), **(B)** Golden (GOLD) vs. Pencil (PENC), and **(C)** Pencil (PENC) vs. Jetblack (JETB). The dark slate gray and deep sky blue colors represent the alternative chromosomes, dotted horizontal black line marks the 99th percentile outlier F_{ST} estimated for autosome and Z chromosome. The red arrows identify genes with non-synonymous changes.

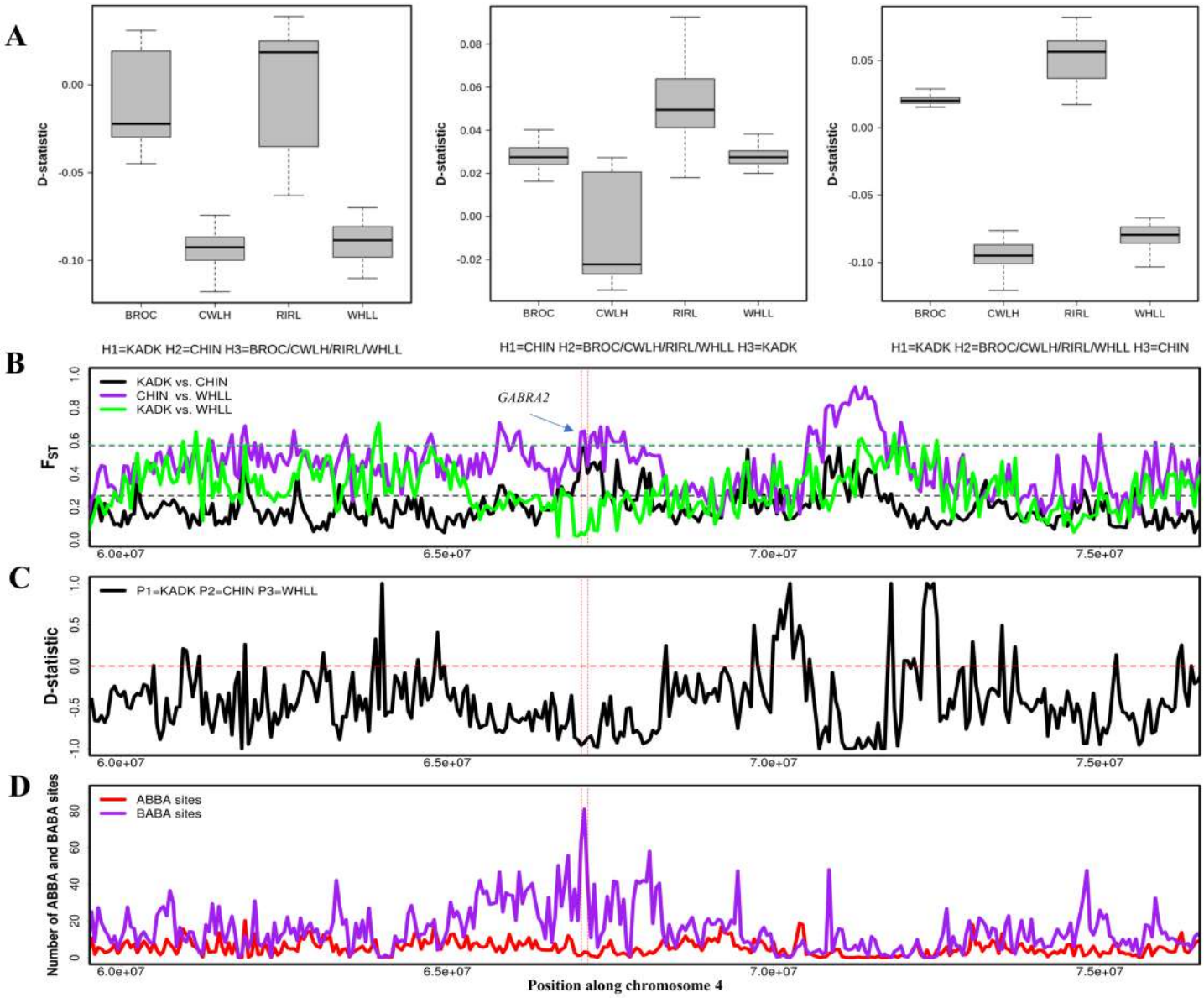
Fig. 5: (A). Genome-wide distribution of Patterson's D-statistic. H1 (KADK) and H2 (CHIN) populations are constant in the first boxplot, H1(CHIN) and H3(KADK) populations in the second, and H1(KADK) and H3(CHIN) populations are constant in the third boxplot. The four commercial populations (BROC/CWLH/RIRL/WHLL) are H3 in the first boxplot and H2 in the second and third boxplot. The red vertical dotted lines represent the candidate introgression regions in all three (B-D) panels. (**B).** Pairwise F_{ST} comparisons between three populations pairs along chromosome 4, where KADK vs. CHIN is a solid black line, CHIN vs. WHLL is a solid purple line, and KADK vs. WHLL is a solid green line. The dotted horizontal lines of the same color represent the 95 percentile outlier for each population comparison. The Blue arrow identifies the gene in the introgressed region (**C).** Patterson's D-statistic along chromosome 4 is shown as a solid black line, and the dotted horizontal red line represents the 0 value. We considered P1 as KADK, P2 as CHIN, P3 as WHLL, and O as RJF population. (**D).** The distribution of ABBA and BABA sites in 50Kb non-overlapping windows are shown in red and purple color lines, respectively.


A
GOLD
PENC
JETB
GOPE
CROS
AYCM
SILK
XBCC
YOSK
LCEM
LCJY
LCMC
LCMY
LCTF
TCAB_FM
TCDQ_FM
TCGZ_FM
B
PC3 (4.71 %)
PC1 (12.7 %)
WHLL
BROL
GOLD
GOPE
JETB
PENC
SILK
TIBC
LCYN
RJFC
NATC
RLCC
BROC
DBLC
SBFC
WCBP
LCEM
LCJY
LCMC
LCMY
LCTF
TCAB
TCAB_FM
TCDQ
TCDQ_FM
TCGZ
TCGZ_FM
TCSN
DGLC
CROS
NONB
YOSK
GGMI
GGSI
GDN4
CWLH
RIRL
AYCM
XBBC
C
Black-bone chickens
Commercial chickens
Native chickens
Admixture proportions for K=7
AYCM
SILK
LCMY
LCJY
LCMC
TCAB_FM
TCDQ_FM
TCGZ_FM
LCEM
XBBC
YOSK
LCTF
KADK
CWLH
WHLL
BROC
RIRL
BROL
GDN4
DBLC
SBFC
WCBP
RLCC
NTAC
LCYN
RJFC
TCAB
TCDQ
TCGZ
TCSN
TIBC
DGLC
GGSI
GGMI


500Kb
127Kb
412Kb
170Kb
500Kb
Flank1
Dup1
Int
Dup2
Flank2
A
Kadaknath (G1)
log (coverage)
B
Red jungle fowl (RJF)
log (coverage)
C
34 BBC Vs 67 non-BBC
Differentiation (F_{ST})
10500000
11000000
11500000
12000000
Position along chromosome


A
KADK vs. CHIN
Differentiation (F_{ST})
0.8
0.6
0.4
0.2
chr
1
2
3
4
5
6
7
8
9
11
13
15
19
22
27
scaffold
Z
B
mean F_{ST} in 10Kb
0.6
0.4
0.2
0.0
HM13
BPIL
TPX2
TTI1
DNTTIP1
ACOT8
SLC12A5
R1
R2
Dup1
π
KADK
CHIN
0.006
0.004
0.002
0.000
iHS
4
3
2
1
0
-1
10000000
10500000
11000000
C
HM13
BPIL
TPX2
TTI1
DNTTIP1
ACOT8
SLC12A5


A
GOLD vs. JETB
B
GOLD vs. PENC
CDH5
NAE1
Differentiation (FST)
C
PENC vs. JETB
USF3
IPCEF1
SYCE3
0.8
0.6
0.4
0.2
chr
1
2
3
4
5
6
7
8
9
11
13
15
19
22
27
scaffold
Z


A
D-statistic
BROC
CWLH
RIRL
WHLL
H1=KADK H2=CHIN H3=BROC/CWLH/RIRL/WHLL
H1=CHIN H2=BROC/CWLH/RIRL/WHLL H3=KADK
H1=KADK H2=BROC/CWLH/RIRL/WHLL H3=CHIN
B
KADK vs. CHIN
CHIN vs. WHLL
KADK vs. WHLL
GABRA2
FST
C
P1=KADK P2=CHIN P3=WHLL
D-statistic
D
ABBA sites
BABA sites
Number of ABBA and BABA sites
6.0e+07
6.5e+07
7.0e+07
7.5e+07
Position along chromosome 4