Hitchhiking of bactericidal/permeability-increasing protein-like gene with the fibromelanosis locus in Kadaknath black-bone-chicken

Short title: Kadaknath-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, genetic linkage, fibromelanosis, *Fm* locus

Abstract

Black-bone-chicken (BBC) meat is popular for its distinctive taste and texture. A complex chromosomal rearrangement at the fibromelanosis (*Fm*) locus on the 20th chromosome results in increased expression of the endothelin-3 (*EDN3*) gene and is responsible for melanin hyperpigmentation in BBC. We use public long-read sequencing data of the silkie breed to resolve high-confidence haplotypes at the *Fm* locus spanning both Dup1 and Dup2 regions and establish that the *Fm_2* scenario is correct of the three possible scenarios of the complex chromosomal rearrangement. The relationship between Chinese and Korean BBC breeds with Kadaknath native to India is underexplored. Our data from whole-genome re-sequencing establishes that all BBC breeds, including Kadaknath, share the complex chromosomal rearrangement junctions at the fibromelanosis (*Fm*) locus. We also identify two *Fm* locus proximal regions (~70Kb and ~300 Kb) with 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 protein domains. Our results indicate that protein-coding changes in the *BPIL* gene hitchhiked with the *Fm* locus in Kadaknath due to close physical linkage. Identifying this *Fm* locus proximal selective sweep sheds light on the genetic 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, exotic, 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, flavor, and use in traditional medicine[9,10]. The black color of BBC results from melanin deposition throughout the body, i.e., melanin hyperpigmentation or fibromelanosis caused by the *Fm* allele[11]. Bateson and Punnet were the pioneers in identifying the autosomal dominant *Fm* allele[12]. Modern studies found a

chromosomal rearrangement on chromosome 20 is involved in the *Fm* locus[13,14]. The overexpression of the endothelin-3 (*EDN3*) gene located within the *Fm* locus is responsible for hyperpigmentation seen in BBC[13,15,16].

BBC breeds have anti-fatigue and anti-hypoxic abilities, with its meat having antioxidant properties[9,17], high carnosine[18,19], and lower fat and cholesterol content[10,20]. Some BBC breeds also have local adaptations. For instance, the Korean Ogye has improved fetal viability and innate immunity against microbial and viral infections[21]. BBC breeds occur globally and have distinct names, such as Ayam Cemani (Indonesia), Black H'Mong (Vietnam), Tuzo (Argentina), Svarthöna (Sweden)[14,22], Yeonsan Ogye (Korea)[23], and Thai BBC (Thailand)[24]. 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[9,25–28]. It is unclear whether all BBC breeds share the same rearrangement at the *Fm* locus, as independent structural variants can produce similar phenotypes[29]. India has a single breed of BBC, commonly known as Kadaknath[19].

Kadaknath is considered native to the Jhabua, Alirajpur, and Dhar districts of Madhya Pradesh, but its farming has recently spread across India [30,31]. The Kadaknath breed has long been documented as a distinctive indigenous Indian breed and is also called Kali Masi or Karaknath [32]. At least three distinctive phenotypes (i.e., Jetblack, Pencil, and Golden) occur within the Kadaknath breed[33]. 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[33]. 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 [19,30,33–41]. 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 analyses are required to understand adaptations in this breed and its genetic history. The aims of this study are : (1) Perform whole-genome re-sequencing 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 for the *Fm* locus by comparing the chromosomal rearrangement junction. (3) Evaluate how the BBC breeds dispersed to various parts of Southeast Asia. (4) Use population genetic statistics to identify signatures of selection in Kadaknath compared to other BBC 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.

Materials and Methods

Population sampling

The study was approved by the Institutional Ethics Committee (IEC) of the Indian Institute of Science Education and Research, Bhopal, vide reference number IISERB/IEC/Certificate/2018-11/03 dated 8th June 2018. We purchased the meat of six individuals (two individuals each from Jetblack, Pencil, and Golden morphs (**Fig. S1**)) 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 two non-black-bone chicken from the same shop to determine the genetic relationship of Kadaknath with native-village chicken and broiler reared in the same poultry (see **Fig. S1**). 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 13 individuals sampled (see **Table S1**).

To evaluate the relationship of Kadaknath chicken globally, we compared our dataset with publicly available chicken re-sequencing data. High-quality re-sequencing datasets of 88 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/). Out of 88 individuals, 22 are BBC, which we considered for further

analysis. Hence, we analyze a dataset of 101 individuals (88 public + 13 sequenced as part of this study) from different breeds (see **Table S2** for more details)[21,23,27,42–47].

Read mapping, variant calling, and phylogeny

We mapped the paired-end raw reads of 101 individuals to the chicken genome assembly (genome assembly version is Gallus_gallus.GRCg6a) using the BWA[48] (Burrows-Wheeler Aligner) v0.7.17-r1188 read mapper with default parameters. We added the read group information using Picard tools and removed duplicate reads in all 101 individual bam files (https://github.com/broadinstitute/picard). We performed the variant calling using FreeBayes[49]with different quality control flags such as --min-alternate-count -C 10, --min-mapping-quality -m 20, --min-base-quality -q 20, and --min-coverage 10. Similarly, we used the bcftools[50],[51] with mapping quality flags such as --mapping quality -C 50, --min base quality -Q 20, and --min mapping quality -q 20 for robust variant calling. We removed the indels from variant calls using vcftools[52] with the -remove-indels flag and extracted the common SNPs using both SNP callers for retaining reliable SNP calls. The Single Nucleotide Polymorphisms (SNPs) identified by both variant callers (bcftools v1.9 and FreeBayes v1.0.0) were used for subsequent analysis. To identify the effect of genetic variants, we used the snpEff[53]. Common SNPs from both vcf files were used to construct phylogeny for 34 and 101 individuals using SNPhylo[54] v20180901, based on a maximum-likelihood tree with 1000 bootstrap values. We excluded the scaffolds, Z, W, and MT chromosomes in phylogeny analysis. The local phylogeny for Dup1 and Dup2 regions was generated using the vk phylo command (using both NJ and UPGMA methods) implemented in VCF-kit[55]. The Mt genome haplotype median-joining networks were constructed using PopART[56] and SplitsTree[57].

Fm locus junction identification

The *Fm* locus consists of a complex chromosomal rearrangement composed of two different non-paralogous regions (Dup1 (~127 Kb) and Dup2 (~170 Kb)) separated by an intermediate (Int) region. Dup1 and Dup2 regions are both duplicated and are involved in a complex rearrangement consisting of two junctions (**A**) Dup1 + (inverted Dup2) and (**B**) (inverted Dup1) + Dup2. 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 used the makewindows command of bedtools[58] (v2.26.0) to create

1Kb non-overlapping windows along chromosome 20. The number of reads in each 1Kb window was calculated using the bedtools coverage command. 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 Gallus_gallus.GRCg6a genome was narrowed down further using coverage estimates in 1bp windows. Dup1 starts at 20:10766772 and ends at 20:10894151. Dup2 occurs further along the chromosome and starts at 20:11306686 and ends at 20:11477501. 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 Gallus_gallus.GRCg6a genome (**Fig. S2, 3**). 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 Gallus_gallus.GRCg6a assembly. Hence, these reads support a rearrangement that leads to junctions between these two regions. Three of these five reads (SRR6189090.111279, SRR6189090.880702, and SRR6189090.387043) support the junction Dup1 + (Inverted Dup2) (i.e., START-DUP1-END-END-DUP2-START), and the other two reads (SRR6189090.54824 and SRR6189090.56386) support the junction (inverted Dup1) + Dup2 (i.e., 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.

Long-read-based resolution of haplotypes at Dup1 and Dup2 regions

Recently, a new high-coverage multi-platform genomic public dataset for the silkie black-bone chicken became available on the ENA (Europen Nucleotide Archive) as part of the Bioproject# PRJNA805080 (we thank China Agricultural University for publishing this data). With a ~65X (Nanopore) and >660X coverage (PacBio) of the same silkie individual, this dataset is well suited to resolve the haplotypes at the *Fm* locus and identify the exact order of rearrangement. Upon visual inspection of the long-read alignments at the Flank-1-Dup1 and Dup2-Flank2 junctions region, three heterozygous sites were recognized at the Dup1 start region, and Dup2 end region, potentially separating the two haplotypes of Dup1 and Dup2. A detailed blastn-based search of reads from both haplotypes anchored one haplotype to the Flank1 + Dup1 junction and another haplotype to the (inverted Dup2) + Dup1 junction. The Flank1 + Dup1 junction containing haplotype is referred to as Dup1 haplotype-1 (D1H1) and contains the allele T at position 10766895. The (inverted Dup2) + Dup1 junction containing haplotype is referred to as Dup1 haplotype-2 (D2H2) and contains the allele A at position 10766895. We extended these haplotypes toward the end of Dup1 by relying upon overlapping long-reads at haplotype-defining sites. To ensure the robustness of our approach, we required that more than ten sequencing reads should support haplotype-specific alleles at each adjacent site (see **Table S3**).

In the Dup2 region, one haplotype was anchored to the Dup2+ Flank2 junction and another haplotype to the Dup2 + (inverted Dup1) junction. The Dup2 + Flank2 junction containing haplotype is Dup2 haplotype-1 (D2H1) and contains the allele A at position 11476819. The Dup2 + (inverted Dup1) junction containing haplotype is Dup2 haplotype-2 (D2H2) and contains the allele G at position 11476819. We extended these haplotypes toward the start of Dup2 by relying upon overlapping long-reads at haplotype-defining sites. We identified 25 well-resolved, high-confidence haplotype-defining sites spanning from the end of Dup2 to the beginning of Dup2 (see **Table S3**). Haplotype-1 of Dup2 contains reads spanning from the end of the Int region to the start of Flank2, and Haplotype-2 of the Dup2 region spans from the end of the Haplotype-1 of Dup1 (D1H1) region to the beginning of Haplotype-2 of Dup1 (D1H2) region.

Principal component analysis and admixture

Principal components analysis (PCA) was performed using PCAngsd[59] based on genotype likelihood estimates from ANGSD[60] (Analysis of Next Generation Sequencing Data) v0.935. 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 used several flags in ANGSD for population structure analysis as follows: -GL 2, -doMaf 1, -minMapQ 30, -minQ 20, -doGlf 2, and -SNP_pval < 1e-6. Genotype likelihood values from ANGSD were used to identify principal components using PCAngsd and genome-wide admixture proportions using NGSadmix[61]. NGSadmix was run for different values of K from K1 to K10 using each K with 15 iterations with flags -minMaf 0.05 and -minInd as specified. We used the online Clumpak[62] web server to find the best K.

Population genetic analysis

Chinese black-bone breeds have diverged to differing extents from Kadaknath and have limited sample sizes. Hence, we combined the individuals from closely related 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) (see **Table S4**). 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[63] on the bam files to quantify callability. For identifying callable regions from mapped bam files for CHIN and KADK individuals, we used GATK with different flags such as --minMappingQuality 20, --minBaseQuality 20, --minDepth 10, --minDepthForLowMAPQ 20, and --maxFractionOfReadsWithLowMAPQ 20. Furthermore, we filtered 50Kb windows with >0.1 repeat element fraction to rule out the possibility of false positives (see **Fig. S4-6**). We used stringent coverage and quality criteria (-GL 2, -dosaf 1, -baq 1, -C 50, -setMinDepthInd 6, -minInd 3, 4 or 9, -minMapQ 30, -minQ 20, and -doCounts 1) in ANGSD to calculate all population genetic statistics. The list of individuals in each population and the population pairs compared is in **Table S4**. We found that at least 20,000 of the 21,659 50Kb windows had sufficient high-quality data in all populations.

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 (Pearson's ρ= 0.98, p-value < 2.2 e-16) (see **Fig. S7, 8**). 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. We estimate haplotype-based statistics iHS and XP-EHH in the rehh[64] R package with option polarised=FALSE using genotype data phased with SHAPEIT[65].

Results

Whole-genome re-sequencing

We generated >25X coverage whole-genome re-sequencing Illumina data for 13 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), one non-black-bone individual where plumage was black (NONB), and one individual of broiler breed (**Fig. S1** and **Table S1**). We compared these Kadaknath chicken genomes with public re-sequencing datasets of chicken 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 S2,** 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 (**Fig. S9**).

Kadaknath is a distinct breed of black-bone fowl

The major axes of genetic variation in BBC (PC1:18.9% and PC2:13%) separated the Kadaknath (KADK), Yeonsan Ogye (YOSK), and Chinese BBC (see **Fig. 1b** and **Fig. S10**). We observed that all Kadaknath individuals form a single cluster distinct from other black-bone (from Indonesia, China, and South Korea) (**Fig. 1b** and **Fig. S10**). Similarly, in the genome-wide PCA analysis of 101 individuals, Kadaknath forms a separate cluster from other commercial, native, and BBC breeds in different PC comparisons (see **Fig. S11**). Admixture analysis suggests Kadaknath has sub-structure but remains distinct from the other black-bone breeds at best K=7 (see **Fig. 1c** and **Fig. S12, 13**). In Chinese BBC, the admixture analysis identified three groups that support the clustering pattern in PCA, while YOSK forms a separate cluster (see **Fig. 1c** and **Fig. S10**). We observed lower genetic diversity (π = 0.002, θ = 0.002) in YOSK, while Kadaknath and Chinese BBC have comparable genetic

diversity (π = 0.004 and 0.003, θ = 0.003 in both) (**Fig. S14a** and **b**). In contrast to the nuclear genome, the mitochondrial genome haplotype network did not separate Kadaknath from other BBC breeds (**Fig. S15, 16**). The *Fm* locus region on chromosome 20, which codes for the black-bone phenotype, is the defining feature of all BBC. Even after excluding chromosome 20, the PCA of the remaining chromosomes finds that Kadaknath is genetically distinct from other BBC breeds (**Fig. S17**). Hence, the genetic distinctiveness of Kadaknath is spread across the entire genome.

All black-bone chicken breeds share the same rearrangement junctions at *Fm* locus

In the non-black chicken, the Dup1 (~127 Kb), Int (~412 Kb), and Dup2 (~170 Kb) regions occur in a single copy, are arranged sequentially, and are flanked by Flank1 (~500 Kb) and Flank2 (~500 Kb) regions (see **Fig. 2a**). While the Dup1 region contains five protein-coding genes (*EDN3*, *ZNF831*, *SLMO2*, *ATP5E*, *TUBB1*), the Dup2 region probably consists of only long-non-coding RNA genes. All non-black chickens have a single copy of this region, referred to as **N* locus (see **Fig. 2a**). The corresponding locus in black chickens is known as the *Fm* locus. The *Fm* locus consists of three different non-paralogous regions (Dup1, Int, and Dup2) that form a complex chromosomal rearrangement in which both the Dup1 and Dup2 regions are duplicated, giving rise to two junctions (**A**) Dup1 + (inverted Dup2) and (**B**) (inverted Dup1) + Dup2. Although the exact ordering of these regions in the rearrangement is not conclusively established, both regions are known to be duplicated due to these regions having a sequencing coverage that is twice the genomic average [13]. The presence of Dup1 + (inverted Dup2) and (inverted Dup1) + Dup2 junctions has been verified in several black chicken breeds [13,14,23]. Based on this information, three possible scenarios have been proposed by earlier studies (see **Fig. 2b**). The **Fm_2* scenario is supported based on crosses between the black bone and non-black bone chicken [13]. However, both **Fm_2* and **Fm_3* scenarios require two rearrangement events compared to a single rearrangement event needed for the **Fm_1* scenario [23].

Distinguishing between these three (**Fm_1*, **Fm_2*, and **Fm_3*) scenarios requires long-range connectivity information such as long-read sequencing data (PacBio, Nanopore, Synthetic long reads, etc.), Hi-C (high-resolution chromosome conformation capture)[66,67], or Optical Mapping[68]. Both **Fm_1* and **Fm_2* contain the same adjacent regions for Dup1 (defined as Haplotype-1 and Haplotype-2 in **Fig. 2b**) inconsistent with **Fm_3*. Hence, long-range information that can span the entire ~127 Kb Dup1 region in a single read will be able

to distinguish between (**Fm_1* and **Fm_2*) vs. **Fm_3*. However, to differentiate between **Fm_1* and **Fm_2* scenarios, we need to span the entire ~170 Kb Dup2 region in a single read. Spanning the Dup1 or Dup2 regions with PacBio (average read lengths of 10 to 25 Kb) or Nanopore (average read lengths of 10 to 30 Kb) technologies with a single read is challenging but not impossible as reads longer than 1 Mb are possible [69,70]. Without such individual reads that can span the entire region, it is possible to perform read-based phasing to infer haplotypes that extend to distinct junctions [71].

Using a public long-read dataset, we have inferred distinct high-confidence Dup1 and Dup2 haplotypes (supporting **Fm_2* scenario). Dup1 haplotypes are anchored by (Flank-1 end)/(Dup2 start) at the Dup1 start and (Int start)/(Dup2 end) at the Dup1 end (**Fig 3**). Each haplotype can be distinguished at 24 sites (see **Table S3**) along the Dup1 region. Each haplotype-defining site shares more than ten long reads with haplotype-specific alleles at adjacent positions. Haplotype-1 (see **Fig. 3a** and **Fig. S18a**, **b**) spans Dup1 from Flank-1 to haplotype-2 Dup2 (D2H2) end, and haplotype-2 (see **Fig. 3b** and **Fig. S18a**, **b**) spans Dup1 from haplotype-2 Dup2 (D2H2) start to Int start. Our inference of Dup1 haplotype-1 (i.e., Flank-1 + Dup1 + (inverted Dup2)) and haplotype-2 (i.e., (inverted Dup2) + Dup1 + Int) sequences rule out the **Fm_3* scenario, suggesting that either **Fm_1* and **Fm_2* scenario is possible. To verify the correct arrangement, we used the read phasing approach on the Dup2 region, and we could distinguish the two haplotypes at the Dup2 region based on 25 sites (see **Table S3**). Haplotype-1 of the Dup2 region (see **Fig. 4a** and **S19a, b**) spans Dup2 from the end of Int to that start of Flank2, and haplotype-2 (see **Fig. 4b** and **S19a, b**) spans Dup2 from haplotype-2 Dup1 (D1H2) start to haplotype-1 Dup1 (D1H1) end. Our inference of haplotype-1 Of Dup2 (i.e., Int+ Dup2 + Flank2) and haplotype-2 (i.e., start of Dup1 + (inverted Dup2) + Dup1 end) sequences rule out the **Fm_1* and **Fm_3* scenarios, suggesting that **Fm_2* scenario is correct (see **Fig. S20**).

Conclusive inference regarding which scenario is present in BBC has been extremely challenging due to the large (~1Mb) size and complexity of the rearrangement. However, in the case of Kadaknath, it has not even been established whether the Dup1 + (inverted Dup2) and (inverted Dup1) + Dup2 Junctions identified in other BBC are present. We have compared the normalized short-read coverage at these junctions to evaluate whether all black-bone chicken breeds share the same rearrangement junctions at the *Fm* locus. The normalized short-read coverage in black-bone chicken (**Fig. 5a-e**) abruptly increases at Dup1 and Dup2

regions, while no such increase occurs in non-black-bone (**Fig. 5f**) chicken. The drastic change in coverage at Dup1 and Dup2 boundaries occurs at the same base in all black-bone chickens (see **Fig. S21-22a-d**). In the case of non-black-bone chickens, no change in coverage occurs at this position (see **Fig. S23**). We also identified the black-bone-specific *Fm* locus junctions Dup1 + (inverted Dup2) and (inverted Dup1) + Dup2 using published PacBio data (see **Fig. 6a, b,** and **Methods**). Short-read coverage spanning these black-bone-specific *Fm* locus junctions is present in black-bone and missing in non-black-bone chicken (**Fig. S24-26**). We can infer that black-bone chicken defining *Fm* locus originated through a single event based on coverage and rearranged junction sequence.

Crossbreeding with native and commercial breeds affects the genome-wide patterns of genetic variation in black-bone breeds (see **Fig. S27-29**). Hence, we focussed on the *Fm* locus to investigate the history of the BBC breeds. The F_{ST} between black-bone and non-black-bone chicken along chromosome 20 is elevated at the Dup1 and Dup2 regions (see **Fig. S30**). However, other population genetic parameters, such as genetic diversity (π and θ) and divergence (D_{xy}), lack any prominent signatures (**Fig. S31-33**). As evident from the elevated F_{ST}, the major axis of genetic variation (assessed using local PCA) in Dup1 and Dup2 regions separates the BBC from non-black-bone chickens (**Fig. S34**). A phylogenetic tree of the SNPs from Dup1 and Dup2 regions also largely separates the BBC breeds from non-BBC breeds (**Fig. S35a** and **35b**). Our evaluation of the genetic differentiation (F_{ST}) landscape between BBC breeds found reduced F_{ST} at Dup1 and Dup2 regions compared to the genomic background (see **Fig. S36**). However, comparing individual BBC breeds with non-black chicken breeds showed the opposite pattern with elevated F_{ST} at Dup1 and Dup2 regions (see **Fig. S37, 38**). Hence, the patterns of genetic differentiation (F_{ST}), local PCA, and phylogenetic tree also support a common origin of the *Fm* locus in all BBC breeds.

Isolation by distance pattern suggests dispersal between India and China

Our analysis discovered that the rearrangement junctions in all BBC are identical to the one in Kadaknath and strongly support a common origin for the *Fm* locus. An independent origin for the *Fm* locus would mean separate rearrangement events have created identical junction sequences. Given the lack of repeat sequences at the junctions, such independent origins seem highly unlikely and are not consistent with the genetic relatedness of BBC and non-BBC breeds at the *Fm* locus. Hence, the current distribution of BBC breeds across Southeast Asia needs an explanation. We hypothesized (see a schematic of the proposed scenario in

Fig. S39) that the dispersal of BBC occurred after a common origin for the *Fm* locus, followed by recent crossbreeding with native and commercial breeds.

To test our hypothesis and delineate the dispersal route 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 **Fig. S40**, mantel's r=0.64, p-value=0.0002). The IBD pattern persisted even after we repeated the analysis using ANGSD/variant-call-based estimates of nucleotide differentiation (F_{ST}) after excluding various populations to avoid errors due to auto-correlation and biased estimates from isolated populations (see **Fig. S41-47**). Among the black-bone breeds, Tibetan black-bone (TBTC) and Sichuan black-bone (LCTMJ) chickens are genetically closest to Kadaknath (mean F_{ST}~0.17). The other black-bone breeds that are geographically more distant from Jhabua (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 S5**). The pattern of IBD spanning India and China suggests potentially human-mediated dispersal. We lack conclusive evidence to identify the direction of dispersal. However, analysis of our dataset, even after excluding alleles found in dbSNP[72] (**Fig. S48, 49**), found that Kadaknath has almost twice the number of private alleles than Chinese and Korean BBC breeds and suggests an India-to-China dispersal. More extensive fine-scale sampling may provide a definitive answer regarding the direction and timing of the dispersal.

Genome-wide signatures of selection in Kadaknath chicken

We screened the genome for selection signatures to identify all Kadaknath-specific regions by comparing Kadaknath with Chinese BBC. We also compared Kadaknath and Chinese BBC (XBBC) with the Korean Yeonsan Ogye (YOSK) to identify population-specific signatures. 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. 7a** and **Table S6**). A pairwise comparison between KADK and CHIN revealed 137 genic regions in the top 1% F_{ST} windows (see **Table S7** and **Fig. S50-82**). We further shortlisted candidates using the additional criteria that genetic diversity (π and θ) and integrated haplotype score (iHS) should be strikingly different between KADK and CHIN. Two prominent signatures from these shortlisted regions indicative 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 ~300 Kb is ~160 Kb before Dup1, and the second region (R2) of ~70 Kb is ~0.7 Mb before 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. 7b** and **Fig. S69**). The elevated iHS in KADK compared to CHIN and the positive value of extended haplotype homozygosity (XP-EHH) supports a selective sweep in KADK. This region also occurs in the top 10% genome-wide genetic divergence (D_{xy}) windows (see **Fig. S83**).

In the first region (R1), 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 first sweep region (R1), also has a non-synonymous change (**Fig. 7c** and **Table S8, 9**). The second sweep region (R2) contains three genes with non-synonymous changes: Histocompatibility Minor 13 (*HM13*), Bactericidal/permeability-increasing protein-like (*BPIL*), and Targeting protein for Xklp2 (*TPX2*). 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 **Fig. S84**). 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.

The Korean Yeonsan Ogye (YOSK) breed is closer to the Chinese XBBC (mean F_{ST} = 0.26) than to Kadakanth (mean F_{ST}=0.3). However, the genome-wide mean F_{ST} is still low enough to distinguish selection signatures. A comparison of YOSK with KADK confirmed that the R1 and R2 regions are specific to Kadaknath (see **Fig. S85** and **Table S6**). Surprisingly, our comparisons with YOSK identified a third region (R3) of ~570 Kb, which is ~2.16 Mb before Dup1 with a sweep specific to YOSK (see **Fig. S86** and **Table S9**). Local PCA of the *Fm* locus, R1, and R2 regions demonstrates that Flank1, R1, and R2 regions have changed only in Kadaknath and not in other BBC (**Fig. S87-89**). A careful examination of the entire chromosome 20 failed to identify any other sweep regions (see **Fig. S90**). Similar to the sweep in KADK, we found two genomic regions with signatures of a sweep in the CHIN population (see **Table S9**). The sweep region on Chr 4 contains the *PCDH7* gene, and the region on Chr 9 contains the *COL4A3* and *MFF* genes (see **Fig. S53, 58**). However, none of these genes have any non-synonymous fixed differences.

Discussion

All BBC breeds share a common origin of the *Fm* locus

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 between 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[14]. The following four lines of evidence support this conclusion: (1) The short-read coverage along the *Fm* locus, (2) the high sequence identity of the rearranged junctions across the BBC breeds, (3) the local and phylogenetic relationship between chicken breeds at the *Fm* locus, and (4) the patterns of pairwise genetic differentiation between BBC and non-BBC breeds.

Spread of BBC across Southeast Asia

The earliest records of black-bone chicken have been found in the writings of Marco Polo, the Venetian explorer-writer who traveled through Asia [13]. The Compendium of Materia Medica or Bencao Gangmu, compiled and edited by Li Shizhen and published in the late 16th century, attributes various medicinal properties to BBC [73]. In Korea, black-bone chicken is thought to have had a royal connection, and "Dongui Bogam", a traditional Korean medical encyclopedia compiled & edited by Heo Jun in 1613, records the medicinal use of BBC [23]. While an earlier study notes that "In 1635 AD, the finding of chickens with black meat (typical of fibromelanosis, FM mutation) in Mozambique suggested direct introductions from India."[2] we could not find literature on how Kadaknath ended up in Jhabua, India.

Jhabua is close to the ancient port cities of Bharuch (260 Km) and Lothal (300 Km) on the west coast of India. Given the proximity to port cities and the prevalence of BBC only in Western India (and the spread of BBC to Africa), we suspected that Kadaknath may have traveled through a marine route. Ancient sea trade between India, Korea, and other parts of Asia are well documented [74]. Moreover, the movement of domesticated breeds during the period of colonialism was also facilitated by the common rule of large parts of Southeast Asia by various European powers [75].

Historical records are patchy, prone to error/obfuscation, and fail to provide conclusive information about the spread of BBC across Asia. Hence, we evaluated the relationship between BBC breeds and patterns of isolation by distance to infer the ancient dispersal route. Although each BBC breed has considerable genetic distinctiveness, population genetic analyses support the common heritage of all BBC breeds and reveal a trend of isolation by

distance (IBD). While IBD patterns are well established in Chinese native chicken breeds[76], 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.

The BBC chicken from Tibet is genetically closest to the Kadaknath breed and suggests that the old Tibet-Nepal salt trade route or the maritime silk 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[25,77]. 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 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[2,47] 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. 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. The beliefs and practices of tribal communities may have contributed to the domestication and conservation of Kadaknath[78].

Conclusive resolution of the complex chromosomal rearrangement

Earlier studies of the *Fm* locus have proposed three possible scenarios for the complex chromosomal rearrangement. Although the **Fm_2* scenario was favored based on crosses between BBC and non-BBC breeds, the correct scenario was not established by previous studies as the genome assembly of this region is challenging. We use a haplotype phasing approach relying upon published long-read datasets to conclusively resolve the correct arrangement to be **Fm_2* scenario at the *Fm* locus. In the **Fm_2* scenario, the distal region (Dup1+Int+Dup2+Flank2) is similar to the **N* arrangement found in non-BBC breeds. However, the proximal region (Flank1+Dup1+(inverted Dup2)) has a very different arrangement than the non-BBC one. In contrast to the **Fm_2* scenario, the proximal region would be similar to the **N* arrangement in **Fm_3* scenario. In the **Fm_1* scenario, only the first occurrence of Dup1 and last occurrence of Dup2 are similar to the **N* arrangement. Hence, in the **Fm_2* and **Fm_3* scenarios recombination with **N* arrangement may be easier (i.e., recombination is supressed only in the inverted ~127Kb (Dup1) or 170Kb (Dup2) region) compared to * *Fm_1* scenario (i.e., recombination is supressed in ~709Kb (inverted (Dup2+Int+Dup1)) region). The ease of recombination with **N* arrangement could explain the prevalence of **Fm_2* scenario which requires two rearrangement events.

Selective sweep near *Fm* locus may be a consequence of linkage

Our analyses 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, which binds and neutralizes lipopolysaccharides (LPS) from the outer membrane of Gram-negative bacteria[79–81]. In the chicken genome, five *BPI*-like genes occur near *BPI*[82], which help arrest bacterial growth, are prominent in phagocytosis

by neutrophils, and work as a bactericidal protein[83,84]. 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[85]. Notably, another BBC breed from Korea, YOSK, has signatures of selection at the TLR4 gene involved in detecting Gram-negative bacteria[21]. Hence, various immune genes may have been selected in different chicken 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[6,86,87]. Hitchhiking of alleles can increase the frequency of both beneficial and mildly deleterious alleles [88–95]. 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 and animals is also known[96–105]. 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 R1 and R2 selective sweeps are found only in KADK, R3 in YOSK, and none in Chinese BBC breeds. Hence, the genetic variants and the selective sweep may represent recent events after the BBC breeds separated from each other. 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.

Genome-wide data generated as part of this study and comparative analysis with other BBC breeds establishes the genetic uniqueness of Kadaknath that extends beyond the *Fm* locus. We also identify specific genes with signatures of selection that are likely to be responsible for the Kadaknath-specific phenotypes. Co-selection of genes that are in linkage, as shown in Kadaknath, is widespread in domestic species[6,86,96–99]. 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.

References

1. 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: 693. doi:10.1038/S41422-020-0349-Y

2. Tixier-Boichard M, Bed'Hom B, Rognon X. Chicken domestication: from archeology to genomics. C R Biol. 2011;334: 197–204. doi:10.1016/J.CRVI.2010.12.012

3. Lawal RA, Hanotte O. Domestic chicken diversity: Origin, distribution, and adaptation. Anim Genet. 2021;52: 385–394. doi:10.1111/AGE.13091

4. Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, et al. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genet. 2008;4. doi:10.1371/JOURNAL.PGEN.1000010

5. Cogburn LA, Porter TE, Duclos MJ, Simon J, Burgess SC, Zhu JJ, et al. Functional genomics of the chicken--a model organism. Poult Sci. 2007;86: 2059–2094. doi:10.1093/PS/86.10.2059

6. Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464: 587–591. doi:10.1038/nature08832

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

8. Mir NA, Rafiq A, Kumar F, Singh V, Shukla V. Determinants of broiler chicken meat quality and factors affecting them: a review. J Food Sci Technol. 2017;54: 2997–3009. doi:10.1007/S13197-017-2789-Z

9. Dou T, Yan S, Liu L, Wang K, Jian Z, Xu Z, 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. 2022;23: 1–15. doi:10.1186/s12864-022-08388-w

10. 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. 2008;87: 160–169. doi:10.3382/PS.2006-00398

11. Muroya S, Tanabe RI, Nakajima I, Chikuni K. Molecular characteristics and site specific distribution of the pigment of the silky fowl. J Vet Med Sci. 2000;62: 391–395. doi:10.1292/JVMS.62.391

12. Bateson BW, -a F, Punnett RC, Contents M. The inheritance of the peculiar pigmentation of the silky fowl. J Genet 1911 13. 1911;1: 185–203. doi:10.1007/BF02981551

13. Dorshorst B, Molin AM, Rubin CJ, Johansson AM, Strömstedt L, Pham MH, et al. A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 2011;7.

14. Dharmayanthi AB, Terai Y, Sulandari S, Zein MSA, Akiyama T, Satta Y. The origin and evolution of fibromelanosis in domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds. PLoS One. 2017;12. doi:10.1371/JOURNAL.PONE.0173147

15. Shinomiya A, Kayashima Y, Kinoshita K, Mizutani M, Namikawa T, Matsuda Y, et al. Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (fibromelanosis) in silky chickens. Genetics. 2012.

16. Zhang P, Cao Y, Fu Y, Zhu H, Xu S, Zhang Y, et al. Revealing the Regulatory Mechanism of lncRNA-LMEP on Melanin Deposition Based on High-Throughput Sequencing in Xichuan Chicken Skin. Genes (Basel). 2022;13: 2143. doi:10.3390/GENES13112143/S1

17. 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. 2009;114: 1345–1350. doi:10.1016/J.FOODCHEM.2008.11.015

18. Tian Y, Xie M, Wang W, Wu H, Fu Z, Lin L. Determination of carnosine in Black-Bone Silky Fowl (Gallus gallus domesticus Brisson) and common chicken by HPLC. Eur Food Res Technol. 2007;226: 311–314. doi:10.1007/S00217-006-0528-1/FIGURES/1

19. Sharma R, Sehrawat R, Ahlawat S, Sharma V, Parmar A, Thakur MS, 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. 2022;12: 1–12. doi:10.1038/s41598-022-07575-9

20. Tian Y, Zhu S, Xie M, Wang W, Wu H, Gong D. Composition of fatty acids in the muscle of black-bone silky chicken (Gallus gellus demesticus brissen) and its bioactivity in mice. Food Chem. 2011;126: 479–483. doi:10.1016/J.FOODCHEM.2010.11.024

21. Cho Y, Kim JY, Kim N. Comparative genomics and selection analysis of Yeonsan Ogye black chicken with whole-genome sequencing. Genomics. 2022;114. doi:10.1016/J.YGENO.2022.110298

22. Johansson AM, Nelson RM. Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds. Front Genet. 2015;6: 44. doi:10.3389/fgene.2015.00044

23. Sohn J Il, Nam K, Hong H, Kim JM, Lim D, Lee KT, et al. Whole genome and transcriptome maps of the entirely black native Korean chicken breed Yeonsan Ogye. Gigascience. 2018;7: 1–14. doi:10.1093/GIGASCIENCE/GIY086

24. Buranawit K, Chailungka C, Wongsunsri C, Laenoi W. Phenotypic characterization of Thai native black-bone chickens indigenous to northern Thailand. Thai J Vet Med. 2016;46: 547–554.

25. Zhu WQ, Li HF, Wang JY, Shu JT, Zhu CH, Song WT, et al. Molecular genetic diversity and maternal origin of Chinese black-bone chicken breeds. Genet Mol Res. 2014;13: 3275–3282. doi:10.4238/2014.APRIL.29.5

26. Kriangwanich W, Piboon P, Sakorn W, Buddhachat K, Kochagul V, Pringproa K, et

al. Consistency of dark skeletal muscles in Thai native black-bone chickens (Gallus gallus domesticus). PeerJ. 2021;9. doi:10.7717/PEERJ.10728

27. Li D, Sun G, Zhang M, Cao Y, Zhang C, Fu Y, et al. Breeding history and candidate genes responsible for black skin of Xichuan black-bone chicken. BMC Genomics. 2020;21. doi:10.1186/S12864-020-06900-8

28. Huang X, Otecko NO, Peng M, Weng Z, Li W, Chen J, et al. Genome-wide genetic structure and selection signatures for color in 10 traditional Chinese yellow-feathered chicken breeds. BMC Genomics. 2020;21. doi:10.1186/S12864-020-6736-4

29. Wang Z, Qu L, Yao J, Yang X, Li G, Zhang Y, et al. An EAV-HP Insertion in 5′ Flanking Region of SLCO1B3 Causes Blue Eggshell in the Chicken. PLOS Genet. 2013;9: e1003183. doi:10.1371/JOURNAL.PGEN.1003183

30. Jadhao RG, Kumbhar DD, Wakode S. Understanding the histology and hematobiochemical aspects of Kadaknath an Indian black fowl with an appealing taste native to Dhar district Madhya Pradesh. Br Poult Sci. 2022 [cited 22 Dec 2022]. doi:10.1080/00071668.2022.2154637

31. Tripathi SB, Jatav G, Malik AA, Joshi S, Singh VK, Negi MS, et al. AFLP markers based genetic diversity and population structure analysis of Kadaknath: an indigenous black meat poultry breed of India. Anim Biotechnol. 2022;33: 1045–1055. doi:10.1080/10495398.2020.1865390

32. Slater AE. The Poultry Industry in India. Worlds Poult Sci J. 1945;1: 46–51. doi:10.1079/WPS19450020

33. Haunshi S, Prince LLL. Kadaknath: a popular native chicken breed of India with unique black colour characteristics. https://doi.org/101080/0043933920211897918. 2021;77: 427–440. doi:10.1080/00439339.2021.1897918

34. Sehrawat R, Sharma R, Ahlawat S, Sharma V, Thakur MS, Kaur M, et al. First report on better functional property of black chicken meat from India. Indian J Anim Res. 2021;55: 727–733. doi:10.18805/IJAR.B-4014

35. Haunshi S, Niranjan M, Shanmugam M, Padhi MK, Reddy MR, Sunitha R, et al. Characterization of two Indian native chicken breeds for production, egg and semen quality, and welfare traits. Poult Sci. 2011;90: 314–320. doi:10.3382/PS.2010-01013

36. Sahu Senior Scientist N, Vigyan Kendra K, Virendra Kumar Junior I, Sahu N, Kumar V, Dhruw S, et al. Woman empowerment and uplifting of tribal farmers through Kadaknath poultry farming in Dantewada district of Chhattisgarh. ~ 315 ~ J Entomol Zool Stud. 2019;7: 315–319.

37. Thakur MS, Parmar SS, Kumar A, Tomar SS, Sarkhel BC, Das G, et al. Evaluation of Kadaknath Chicken for Coccidial Resistance by Oocyst Count, Lesion Scoring and Oocyst Index in Eimeria Tenella Infection . J Anim Res. 2015;5: 579. doi:10.5958/2277-940X.2015.00097.2

38. Jena P, Panigrahi B, Panda N, Mohapatra L, Mallik B, Bagh J. Reproductive Performance and Egg Quality Traits of Kadaknath in Intensive Managemental Condition under Hot and Humid Climate. Int J Livest Res. 2018; 1. doi:10.5455/IJLR.20180131054034

39. Shanmathy M, Tyagi JS, Gopi M, Mohan J, Beulah P, Ravi Kumar D. Comparative

Assessment on Performance of Aseel and Kadaknath in Hot and Humid Conditions in Tropics. Int J Curr Microbiol Appl Sci. 2018;7: 2156–2165. doi:10.20546/ijcmas.2018.705.251

40. Rout PK, Pani PK, Naithani S. Genetic susceptibility of indigenous chicks to subgroup A Rous sarcoma virus inoculated via the chorioallantoic membrane. Vet Immunol Immunopathol. 1992;33: 89–102. doi:10.1016/0165-2427(92)90037-Q

41. Haunshi S, Devatkal S, Prince LLL, Ullengala R, Ramasamy K, Chatterjee R. Carcass Characteristics, Meat Quality and Nutritional Composition of Kadaknath, a Native Chicken Breed of India. Foods (Basel, Switzerland). 2022;11: 3603. doi:10.3390/FOODS11223603

42. 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: 1880–1889. doi:10.1093/MOLBEV/MSV071

43. Ulfah M, Kawahara-Miki R, Farajalllah A, Muladno M, Dorshorst B, Martin A, et al. Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. BMC Genomics. 2016;17. doi:10.1186/S12864-016-2652-Z

44. Roux PF, Boutin M, Désert C, Djari A, Esquerré D, Klopp C, et al. Re-sequencing data for refining candidate genes and polymorphisms in QTL regions affecting adiposity in chicken. PLoS One. 2014;9. doi:10.1371/JOURNAL.PONE.0111299

45. Qanbari S, Rubin CJ, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation in modern chicken. PLoS Genet. 2019;15. doi:10.1371/JOURNAL.PGEN.1007989

46. Fan WL, Ng CS, Chen CF, Lu MYJ, Chen YH, Liu CJ, et al. Genome-wide patterns of genetic variation in two domestic chickens. Genome Biol Evol. 2013;5: 1376–1392. doi:10.1093/GBE/EVT097

47. Mariadassou M, Suez M, Sathyakumar S, Vignal A, Arca M, Nicolas P, et al. Unraveling the history of the genus Gallus through whole genome sequencing. Mol Phylogenet Evol. 2021;158. doi:10.1016/J.YMPEV.2020.107044

48. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

49. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012.

50. Li H, Barrett J. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27: 2987–2993. doi:10.1093/BIOINFORMATICS/BTR509

51. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10: 1–4. doi:10.1093/GIGASCIENCE/GIAB008

52. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. doi:10.1093/BIOINFORMATICS/BTR330

53. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, 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). 2012;6: 80–92. doi:10.4161/FLY.19695/SUPPL_FILE/KFLY_A_10919695_SM0001.ZIP

54. Lee TH, Guo H, Wang X, Kim C, Paterson AH. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics. 2014;15: 162. doi:10.1186/1471-2164-15-162

55. Cook DE, Andersen EC. VCF-kit: assorted utilities for the variant call format. Bioinformatics. 2017;33: 1581–1582. doi:10.1093/BIOINFORMATICS/BTX011

56. Leigh JW, Bryant D. popart: full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6: 1110–1116. doi:10.1111/2041-210X.12410

57. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998;14: 68–73. doi:10.1093/BIOINFORMATICS/14.1.68

58. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–2. Available: http://bioinformatics.oxfordjournals.org/content/26/6/841.short

59. Meisner J, Albrechtsen A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics. 2018;210: 719–731. doi:10.1534/GENETICS.118.301336

60. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014;15: 356.

61. Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. Genetics. 2013;195: 693–702. doi:10.1534/GENETICS.113.154138

62. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 2015;15: 1179–1191. doi:10.1111/1755-0998.12387

63. 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: 1297–1303. doi:10.1101/gr.107524.110

64. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 2012;28: 1176–1177. doi:10.1093/BIOINFORMATICS/BTS115

65. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods 2011 92. 2011;9: 179–181. doi:10.1038/nmeth.1785

66. Kronenberg ZN, Rhie A, Koren S, Concepcion GT, Peluso P, Munson KM, et al. Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C. Nat Commun. 2021;12. doi:10.1038/S41467-020-20536-Y

67. Wang X, Luan Y, Yue F. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. Sci Adv. 2022;8.

doi:10.1126/SCIADV.ABN9215

68. Weissensteiner MH, Bunikis I, Catalán A, Francoijs KJ, Knief U, Heim W, et al. Discovery and population genomics of structural variation in a songbird genus. Nat Commun 2020 111. 2020;11: 1–11. doi:10.1038/s41467-020-17195-4

69. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020 211. 2020;21: 1–16. doi:10.1186/S13059-020-1935-5

70. Hon T, Mars K, Young G, Tsai YC, Karalius JW, Landolin JM, et al. Highly accurate long-read HiFi sequencing data for five complex genomes. Sci Data 2020 71. 2020;7: 1–11. doi:10.1038/s41597-020-00743-4

71. Patterson MD, Marschall T, Pisanti N, Van Iersel L, Stougie L, Klau GW, et al. WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. J Comput Biol. 2015;22: 498–509. doi:10.1089/CMB.2014.0157

72. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29: 308–311. doi:10.1093/NAR/29.1.308

73. Wang H, Cahaner A, Lou L, Zhang L, Ge Y, Li Q, et al. Genetics and breeding of a black-bone and blue eggshell chicken line. 1. Body weight, skin color, and their combined selection. Poult Sci. 2021;100. doi:10.1016/J.PSJ.2021.101035

74. Acri A. Maritime Buddhism. Oxford Res Encycl Relig. 2018 [cited 22 Dec 2022]. doi:10.1093/ACREFORE/9780199340378.013.638

75. Sudrajad P, Subiharta S, Adinata Y, Lathifah A, Lee JH, Lenstra JA, et al. An insight into the evolutionary history of Indonesian cattle assessed by whole genome data analysis. PLoS One. 2020;15: e0241038. doi:10.1371/JOURNAL.PONE.0241038

76. Nie C, Almeida P, Jia Y, Bao H, Ning Z, Qu L, et al. Genome-Wide Single-Nucleotide Polymorphism Data Unveil Admixture of Chinese Indigenous Chicken Breeds with Commercial Breeds. Genome Biol Evol. 2019;11: 1847–1856. doi:10.1093/GBE/EVZ128

77. Zhang T, Du W, Lu H, Wang L. Genetic Diversity of Mitochondrial DNA of Chinese Black-bone Chicken. Brazilian J Poult Sci. 2018;20: 565–572. doi:10.1590/1806-9061-2018-0739

78. Dar JA, Subashree K, Raha D, Kumar A, Khare PK, Khan ML. Tree diversity, biomass and carbon storage in sacred groves of Central India. Environ Sci Pollut Res 2019 2636. 2019;26: 37212–37227. doi:10.1007/S11356-019-06854-9

79. Sun YY, 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. 2016;11. doi:10.1371/JOURNAL.PONE.0154045

80. Holweg A, Schnare M, Gessner A. The bactericidal/permeability-increasing protein (BPI) in the innate defence of the lower airways. Biochem Soc Trans. 2011;39: 1045–1050. doi:10.1042/BST0391045

81. Monson MS, Van Goor AG, Persia ME, Rothschild MF, Schmidt CJ, Lamont SJ.

Genetic lines respond uniquely within the chicken thymic transcriptome to acute heat stress and low dose lipopolysaccharide. Sci Rep. 2019;9. doi:10.1038/S41598-019-50051-0

82. Chiang SC, Veldhuizen EJA, Barnes FA, Craven CJ, Haagsman HP, Bingle CD. Identification and characterisation of the BPI/LBP/PLUNC-like gene repertoire in chickens reveals the absence of a LBP gene. Dev Comp Immunol. 2011;35: 285–295. doi:10.1016/J.DCI.2010.09.013

83. Elsbach P, Weiss J. Role of the bactericidal/permeability-increasing protein in host defence. Curr Opin Immunol. 1998;10: 45–49. doi:10.1016/S0952-7915(98)80030-7

84. 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. 2003;31: 785–790. doi:10.1042/BST0310785

85. Sharma A, Gupta S, Patil AB, Vijay N. Birth and death in terminal complement pathway. Mol Immunol. 2022;149: 174–187. doi:10.1016/J.MOLIMM.2022.07.006

86. Webster MT, Kamgari N, Perloski M, Hoeppner MP, Axelsson E, Hedhammar Å, et al. Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. BMC Genomics. 2015;16: 1–17. doi:10.1186/S12864-015-1702-2/FIGURES/4

87. Feder JL, Nosil P, Flaxman SM. Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations. Front Genet. 2014;5: 295. doi:10.3389/fgene.2014.00295

88. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics. 2006;173: 419–34. doi:10.1534/genetics.105.047985

89. Flaxman SM, Feder JL, Nosil P. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. Evolution. 2013;67: 2577–91. doi:10.1111/evo.12055

90. Martin SH, Singh KS, Gordon IJ, Omufwoko KS, Collins S, Warren IA, et al. Whole-chromosome hitchhiking driven by a male-killing endosymbiont. PLoS Biol. 2020;18. doi:10.1371/JOURNAL.PBIO.3000610

91. Huff CD, Witherspoon DJ, Zhang Y, Gatenbee C, Denson LA, Kugathasan S, et al. Crohn's Disease and Genetic Hitchhiking at IBD5. Mol Biol Evol. 2012;29: 101. doi:10.1093/MOLBEV/MSR151

92. Weist P, Jentoft S, Tørresen OK, Schade FM, Pampoulie C, Krumme U, et al. The role of genomic signatures of directional selection and demographic history in the population structure of a marine teleost with high gene flow. Ecol Evol. 2022;12. doi:10.1002/ECE3.9602

93. Hale MC, Campbell MA, McKinney GJ. A candidate chromosome inversion in Arctic charr (Salvelinus alpinus) identified by population genetic analysis techniques. G3 (Bethesda). 2021;11. doi:10.1093/G3JOURNAL/JKAB267

94. Franssen SU, Nolte V, Tobler R, Schlotterer C. Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental Drosophila melanogaster populations. Mol Biol Evol. 2015;32: 495–509. doi:10.1093/MOLBEV/MSU320

95. Bierne N. The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. Evolution. 2010;64: 3254–3272. doi:10.1111/J.1558-5646.2010.01050.X

96. Li J, Zhang Y, Ma R, Huang W, Hou J, Fang C, et al. Identification of ST1 reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. Plant Biotechnol J. 2022;20: 1110–1121. doi:10.1111/PBI.13791

97. Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol. 2017;18. doi:10.1186/S13059-017-1289-9

98. Zhang D, Liu C, Cheng H, Kan G, Cui S, Meng Q, et al. Quantitative trait loci associated with soybean tolerance to low phosphorus stress based on flower and pod abscission. Plant Breed. 2010;129: 243–249. doi:10.1111/J.1439-0523.2009.01682.X

99. Kong L, Lu S, Wang Y, Fang C, Wang F, Nan H, et al. Quantitative Trait Locus Mapping of Flowering Time and Maturity in Soybean Using Next-Generation Sequencing-Based Analysis. Front Plant Sci. 2018;9. doi:10.3389/FPLS.2018.00995

100. Guan J, Xu Y, Yu Y, Fu J, Ren F, Guo J, et al. Genome structure variation analyses of peach reveal population dynamics and a 1.67 Mb causal inversion for fruit shape. Genome Biol. 2021;22. doi:10.1186/S13059-020-02239-1

101. Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. Mol Ecol. 2020;29: 2535–2549. doi:10.1111/MEC.15428

102. Wang X, Gao L, Jiao C, Stravoravdis S, Hosmani PS, Saha S, et al. Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding. Nat Commun. 2020;11. doi:10.1038/S41467-020-19682-0

103. Jayakodi M, Padmarasu S, Haberer G, Bonthala VS, Gundlach H, Monat C, et al. The barley pan-genome reveals the hidden legacy of mutation breeding. Nature. 2020;588: 284–289. doi:10.1038/S41586-020-2947-8

104. Beissinger TM, Hirsch CN, Vaillancourt B, Deshpande S, Barry K, Robin Buell C, et al. A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. Genetics. 2014;196: 829–840. doi:10.1534/GENETICS.113.160655

105. Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LR, Cristobal MS, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 2012;10. doi:10.1371/JOURNAL.PBIO.1001258

Acknowledgments

Funding:

Council of Scientific and Industrial Research Fellowship (SSS)

University Grants Commission Ph.D. scholarship (AS)

The Department of Biotechnology, Ministry of Science and Technology, India (Grant no. BT/11/IYBA/2018/03) (NV)

Science and Engineering Research Board (Grant no. ECR/2017/001430) (NV)

Author contributions:

Conceptualization: SSS, AS, NV

Methodology: SSS, AS

Investigation: SSS, AS, NV

Visualization: SSS, AS

Supervision: NV

Writing—original draft: SSS, AS

Writing—review & editing: SSS, AS, NV

Competing interest statement

None to declare

Data and materials availability:

All primary sequencing data are available from the SRA under BioProject# PRJEB51457 and PRJEB58678. 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**.

Ethics statement

The study was approved by the Institutional Ethics Committee (IEC) of the Indian Institute of Science Education and Research, Bhopal, vide reference number IISERB/IEC/Certificate/2018-11/03 dated 8th June 2018.

Contribution to field

Domestication of chicken occurred in Asia before spreading across the world. Large-scale population genomics has helped unravel the history of domestication and the genetic contribution from the wild relatives of chickens. Yet, different chicken breeds' distinctiveness and adaptation to local environments have received limited attention, especially in India. Hence, population genomic datasets of indigenous breeds from India will help understand the spread of domesticated chicken and breed formation. Some indigenous breeds are experiencing a decline in numbers and need conservation. We focussed on one of the most distinctive indigenous chicken breeds, the Kadaknath chicken, which is well known for its black bone and meat. Commonly known as Kali masi or karaknath, the unique taste and use in traditional medicine have made this breed extremely popular across India.

Black-bone chicken breeds are found in several other countries and have fascinated breeders and geneticists alike. We compared the Kadaknath genome with other black-bone chicken breeds and found that all black-bone breeds share the same rearrangement junctions at the *Fm* locus, responsible for the intense black pigmentation throughout the body. Notably, we use public long-read high-coverage datasets to conclusively resolve the correct arrangement of the complex chromosomal rearrangement at the *Fm* locus.

Figures

Figures

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 34 black-bone chicken individuals. PC1 and PC2 explained 18.9% and 13% variance, respectively. **(c)** Population genetic structure and individual ancestry were estimated using NGSadmix for 34 black-bone chickens from different breeds based on best K=7.

Fig. 2: Arrangement of *Fm* locus region in non-BBC and BBC breeds: Two different non-paralogous regions on chromosome 20 are referred to as Duplication 1 (Dup1) and Duplication 2 (Dup2), shown in Gold and light green colors, respectively. The length of the Dup1 region is ~127 Kb, the Intermediate (Int) is ~412 Kb, and Dup2 is ~170 Kb in size. (**a**) These regions are neither duplicated nor rearranged in nonblack bone chicken (*N). The Dup1, Int, and Dup2 regions are collectively referred to as the *Fm* locus. Dup1 contains five genes, *EDN3*, *ZNF831*, *SLMO2*, *ATP5E*, and *TUBB1*, whereas the Dup2 region does not have any protein-coding genes. (**b**) Three possible scenarios (**Fm_1*, **Fm_2*, and **Fm_3*) for *Fm* locus have been proposed in BBC (earlier described in Dorshorst et al. 2011, Dharmayanthi et al. 2017, and Shon et al. 2018). The dark solid purple line represents haplotype-1, which spans across Dup1 from Flank1 to inverted Dup2 (i.e., Flank-1 + Dup1 + (inverted Dup2)). The blue line represents haplotype-2, which spans Dup1 from inverted Dup2 to Int (i.e., (inverted Dup2) + Dup1 + Int).

Fig. 3: Long-read-based haplotypes resolve the sequence spanning the Dup1 region at *Fm* locus: The two haplotypes spanning the Dup1 region are distinguished by distinct alleles at 24 positions using long sequencing reads. Red dotted vertical boxes highlight the alleles that differ between haplotype-1 and haplotype-2 at the same positions. The alleles at 24 sites

that separate these two haplotypes at various positions along Dup1 are presented sequentially between the red dotted boxes. (**a**) Haplotype-1 of Dup1 spans from Flank1 to inverted Dup2 (i.e., Flank-1 + Dup1 + (inverted Dup2)). The purple lines represent overlapping reads of Nanopore containing Dup1 haplotype-1 alleles. The light blue dotted line represents the span of the same read from the end of haplotype-1 of Dup1 to the end of haplotype-2 of Dup2. (**b**) Haplotype-2 spans Dup1 from haplotype-2 of Dup2 to the Int region (i.e., (inverted Dup2) + Dup1 + Int). The blue lines represent the overlapping reads of Nanopore containing haplotype-2 alleles. The light blue dotted line represents the span of the same read from the start of haplotype-2 of Dup1 to the start of haplotype-2 of Dup2. Both haplotypes of Dup1 are connected to the single haplotype of Dup2 (i.e., haplotype-2 of Dup2) but at different ends.

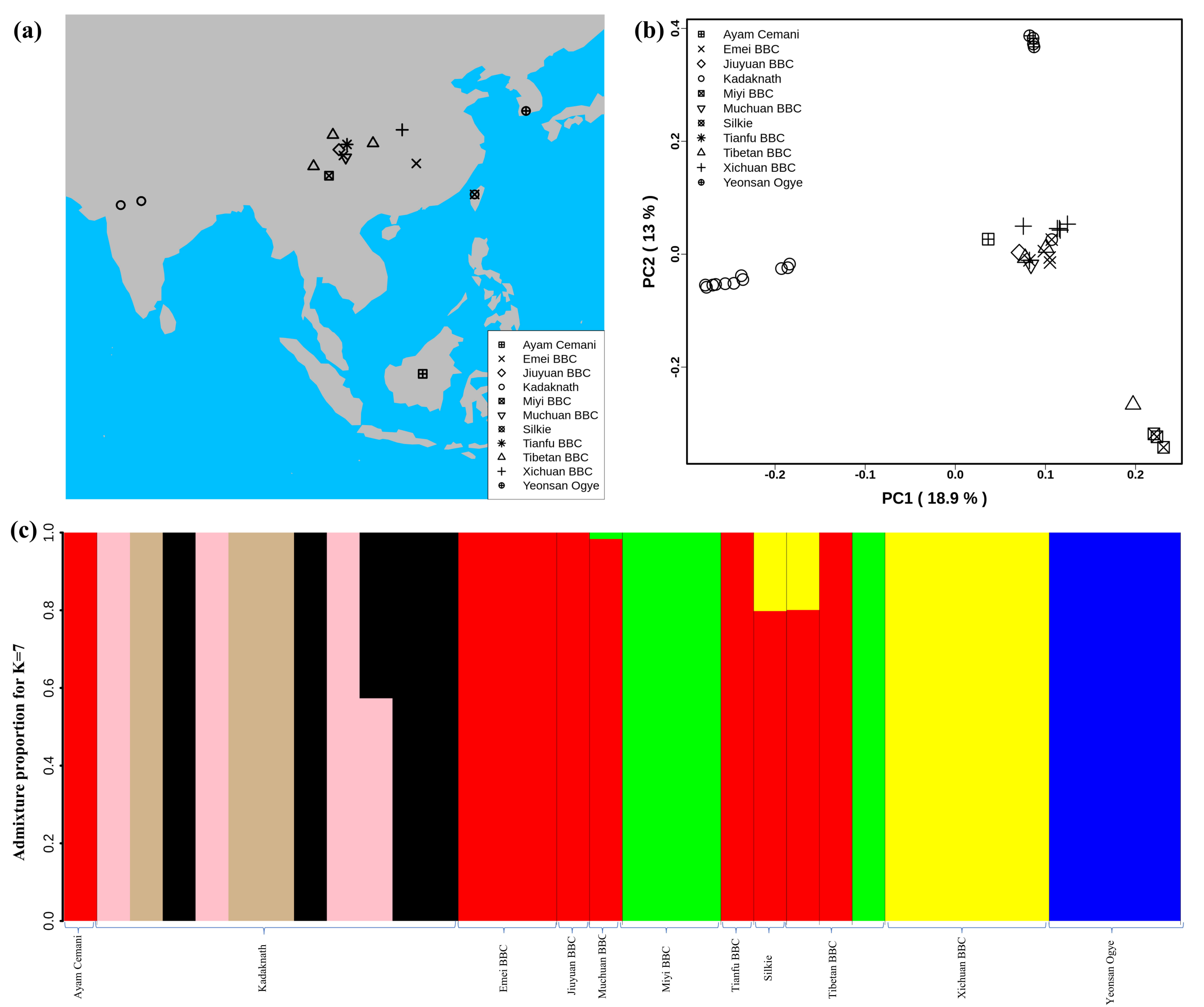
Fig. 4: Long-read-based haplotypes resolve the sequence spanning the Dup2 region at *Fm* locus: The two haplotypes spanning the Dup2 region are distinguished by distinct alleles at 25 positions using long sequencing reads. Red dotted vertical boxes highlight the alleles that differ between haplotype-1 and haplotype-2 at the same positions. The alleles at 25 sites that separate these two haplotypes at various positions along Dup2 are presented sequentially between the red dotted boxes. (**a)** Haplotype-1 of Dup2 spans from Int to Flank2 (i.e., Int + Dup2 + Flank2). The grey lines represent overlapping reads of Nanopore containing Dup2 haplotype-1 alleles. (**b**) Haplotype-2 of Dup2 spans from the start of haplotype-2 of Dup1 to the end of haplotype-1 of Dup1 region (i.e., Dup1 +(inverted Dup2)+ Dup1). The red lines represent the overlapping reads of Nanopore containing Dup2 haplotype-2 alleles. The light blue dotted line represents the span of the same reads from the start of haplotype-2 of Dup1 to the start of haplotype-2 of Dup2. Similarly, the same reads span the end of haplotype-1 of Dup1 to the end of haplotype-2 of Dup2.

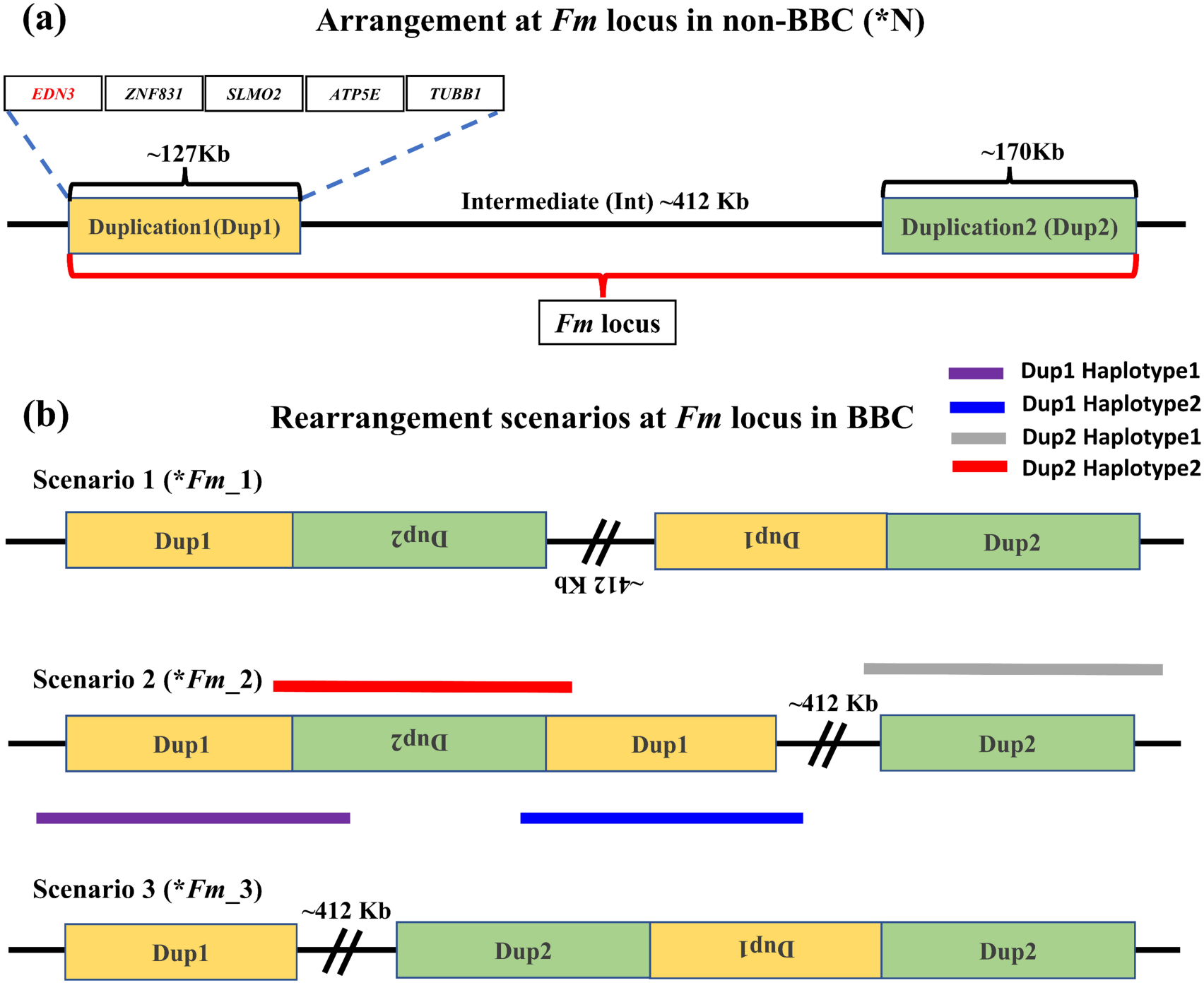
Fig. 5: Comparison of *Fm* locus read coverage in black-bone vs. non-black-bone chicken breeds: Read coverage along chromosome 20 at the *Fm* locus is shown in 1Kb sliding windows. Two duplicated genomic loci, Dup1 and Dup2, are denoted by vertical dotted red lines and have a higher coverage in black-bone chicken breeds: **(a)** Kadaknath, **(b)** Xichuan, **(c)** Yeonsan Ogye, **(d)** Tianfu, **(e)** Ayam Cemani than non-black bone chicken: **(f)** Red junglefowl (RJF).

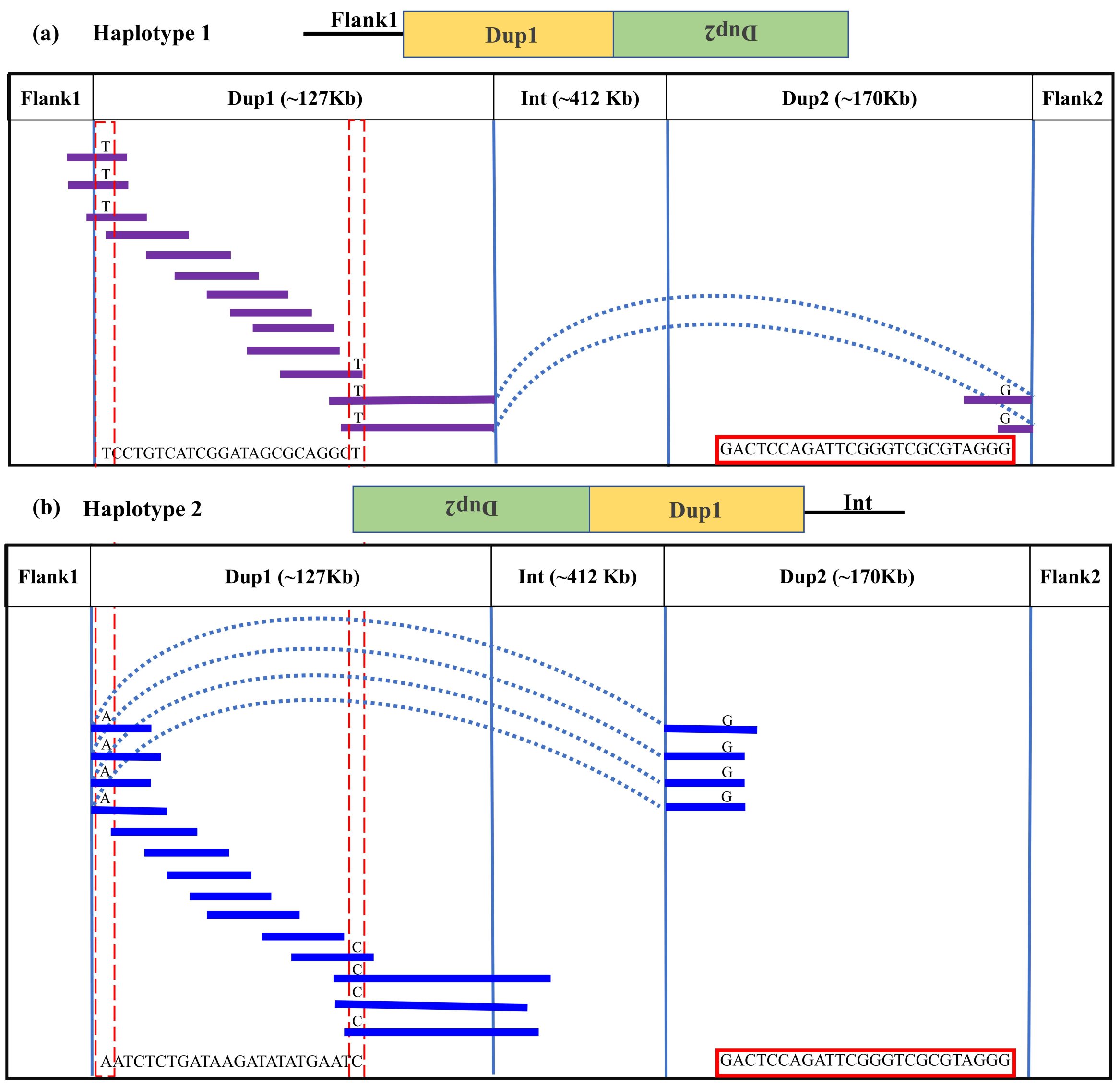
Fig 6: The nucleotide sequence of rearranged junctions at the *Fm* locus in black-bone vs. non-black-bone chicken: (a) Dup1 + (inverted Dup2) **(b)** (Inverted Dup1) + Dup2. Brown arrows represent the direction of rearranged junction sequence, which is present as a

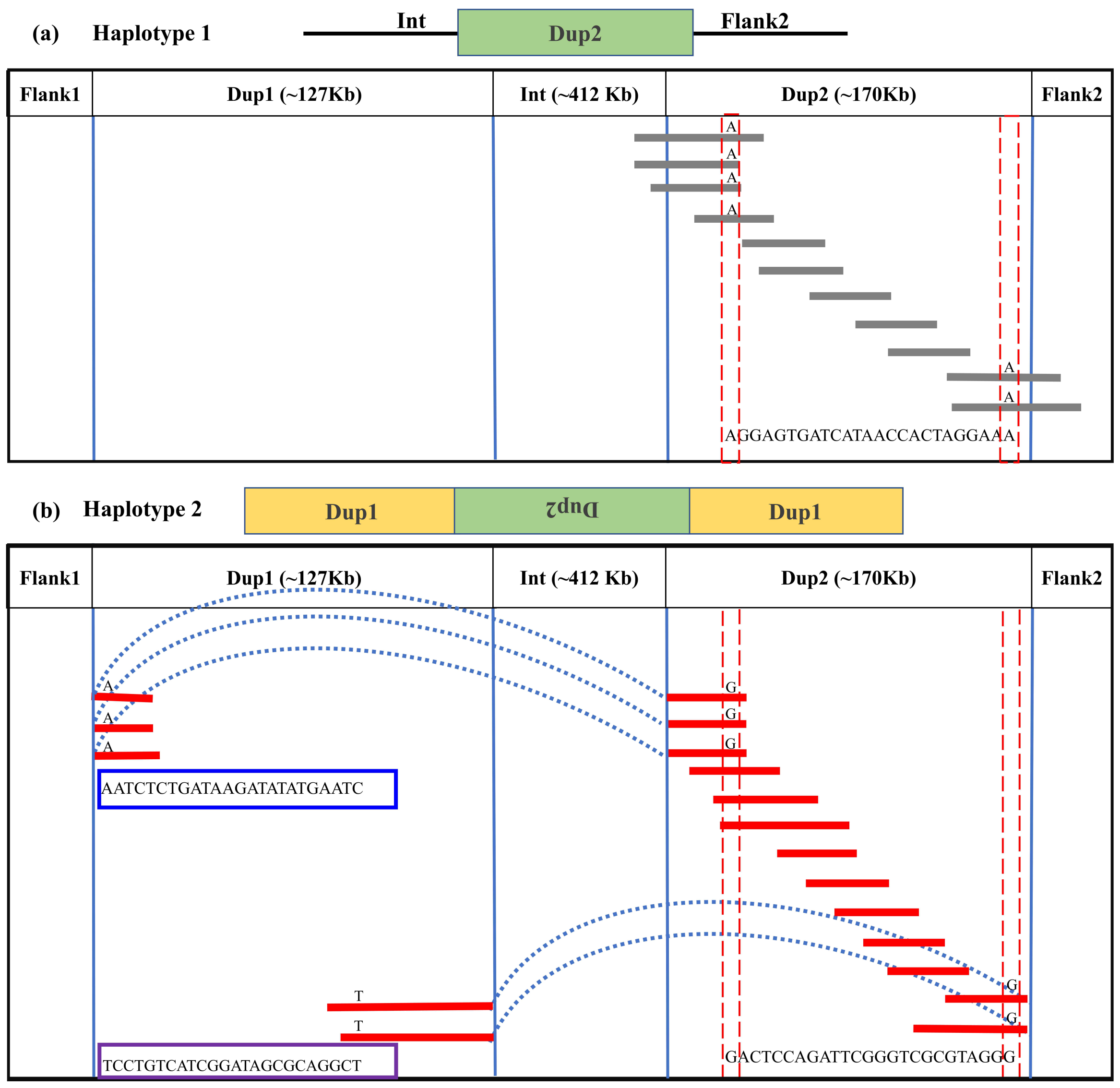
continuous sequence (verified by their presence in a single read) in BBC breeds. In contrast, non-BBC breeds did not contain these junctions. Sequences highlighted in red and blue represent the Dup1 and Dup2 sequences, respectively. Sequences in black color represent the subsequent nucleotides.

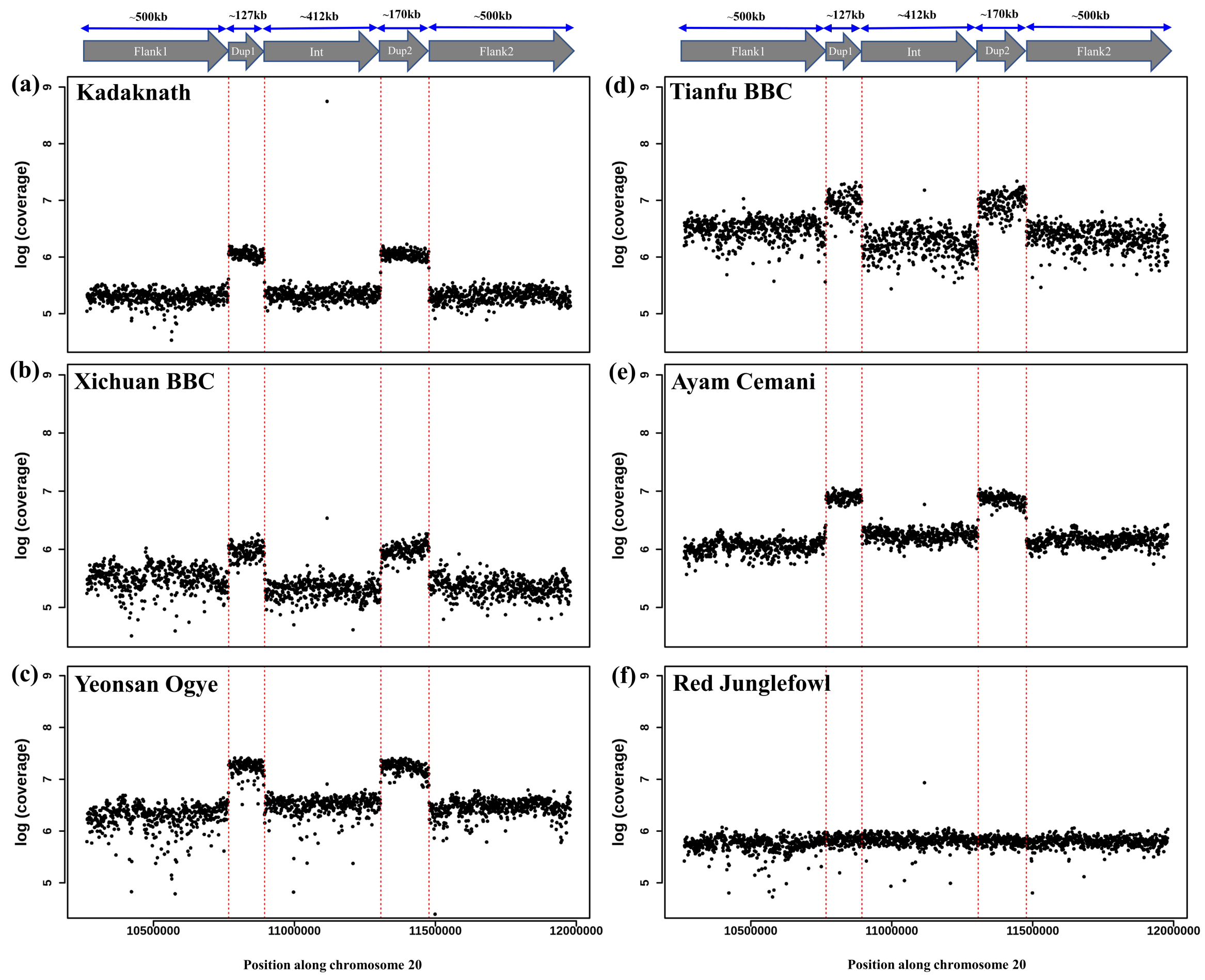
Fig. 7: Within and between population comparison of KADK with CHIN: (a) Genome-wide landscape of pairwise genetic differentiation (F_{ST}) between KADK and CHIN in 50Kb non-overlapping windows. The dark slate gray and deep sky blue colors represent the alternative chromosomes. The dotted horizontal black 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 nucleotide diversity (π) and integrated haplotype score (iHS) with major allele as ancestral and minor allele as derived. The genes are represented as black boxes with names. The horizontal red dotted line marks the 99th percentile 10Kb F_{ST} outlier. 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.

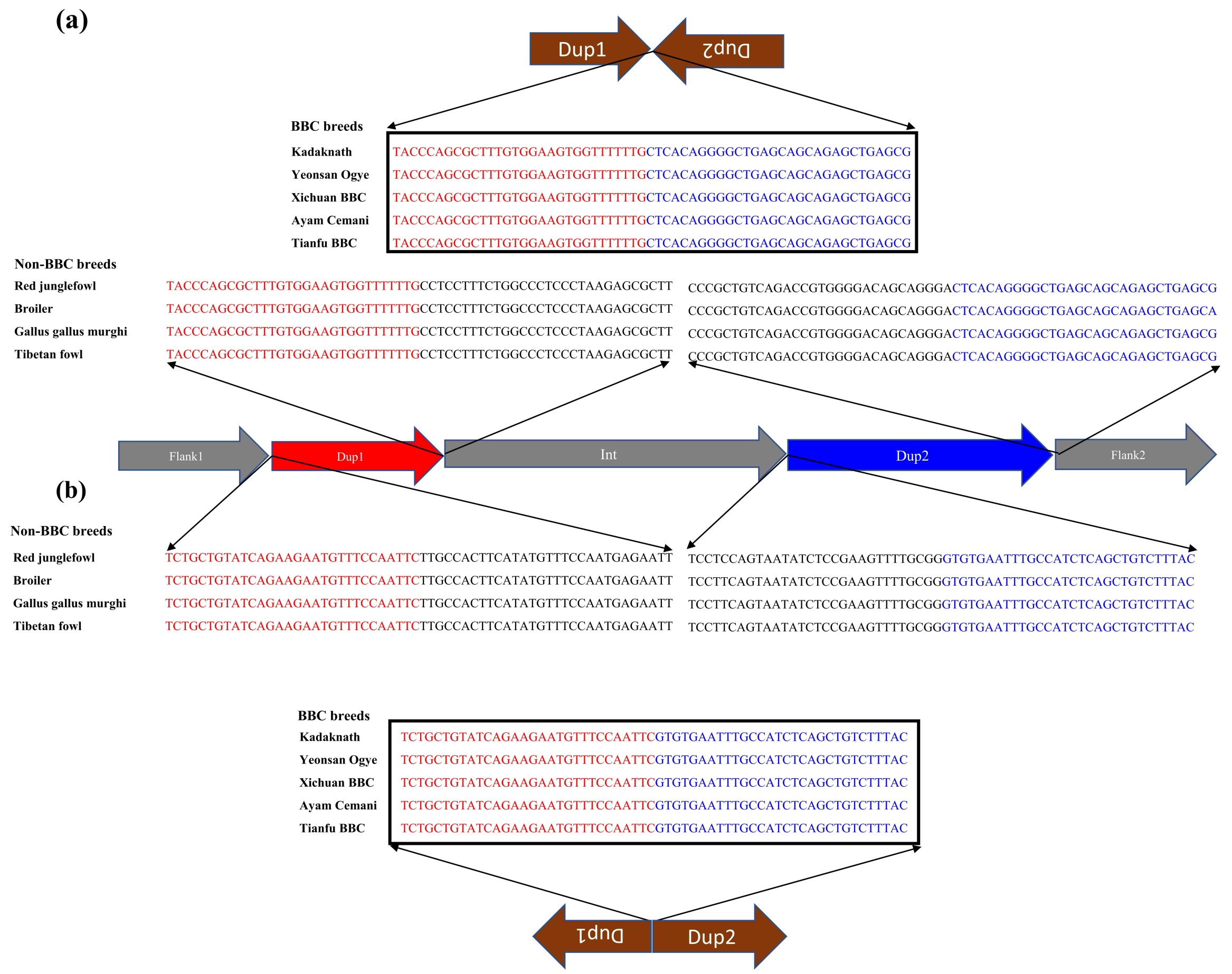

(a)
Ayam Cemani
Emei BBC
Jiuyuan BBC
Kadaknath
Miyi BBC
Muchuan BBC
Silkie
Tianfu BBC
Tibetan BBC
Xichuan BBC
Yeonsan Ogye
(b)
PC2 (13 %)
PC1 (18.9 %)
0.4
0.2
0.0
-0.2
-0.2
-0.1
0.0
0.1
0.2
(c)
Admixture proportion for K=7
1.0
0.8
0.6
0.4
0.2
0.0
Ayam Cemani
Kadaknath
Emei BBC
Jiuyuan BBC
Muchuan BBC
Miyi BBC
Tianfu BBC
Silkie
Tibetan BBC
Xichuan BBC
Yeonsan Ogye

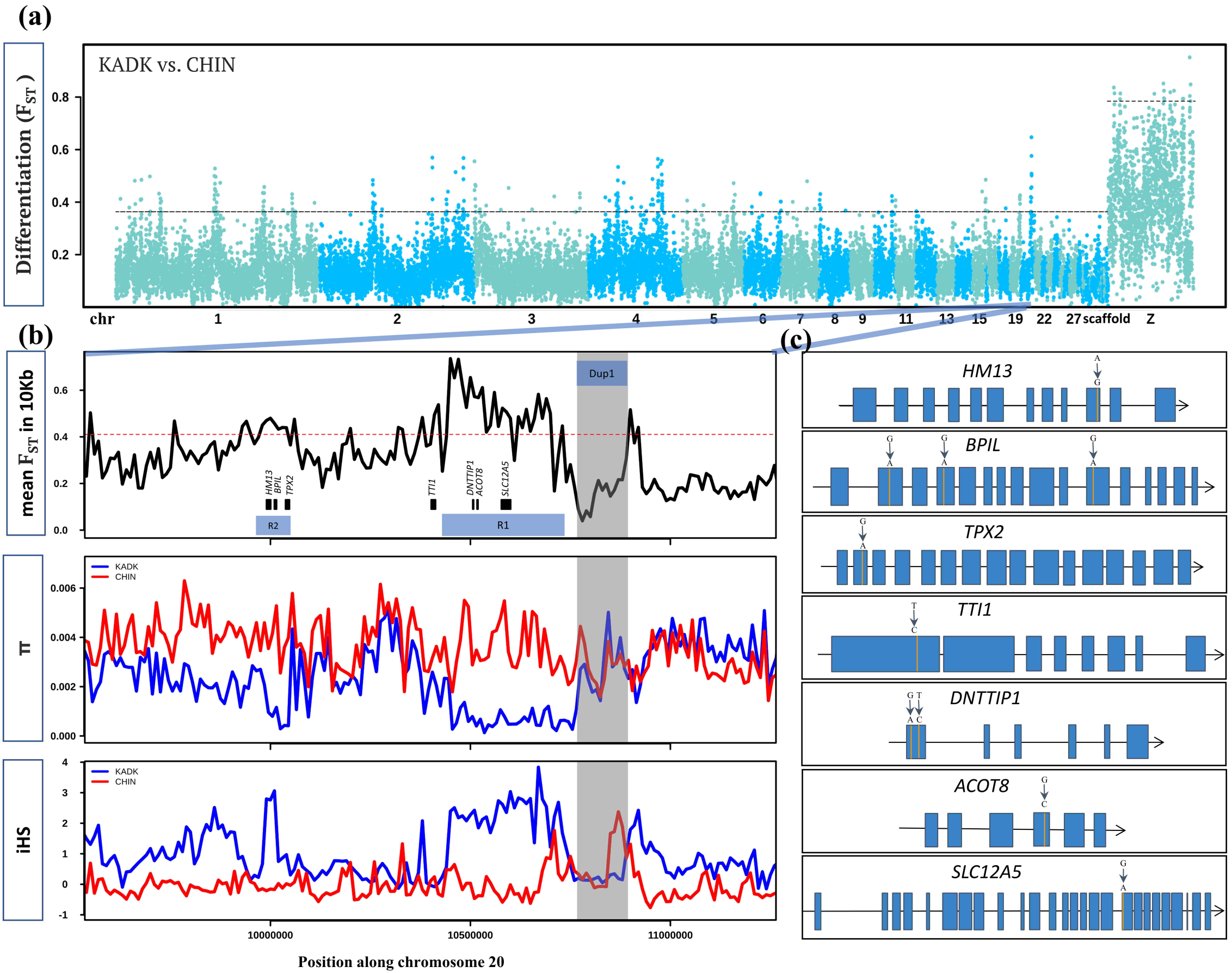

(a)
Arrangement at Fm locus in non-BBC (*N)
EDN3
ZNF831
SLMO2
ATP5E
TUBB1
~127Kb
~170Kb
Intermediate (Int) ~412 Kb
Duplication1(Dup1)
Duplication2 (Dup2)
Fm locus
(b)
Rearrangement scenarios at Fm locus in BBC
Dup1 Haplotype1
Dup1 Haplotype2
Dup2 Haplotype1
Dup2 Haplotype2
Scenario 1 (*Fm_1)
Dup1
Dup2
~412 Kb
Dup1
Dup2
Scenario 2 (*Fm_2)
~412 Kb
Dup1
Dup2
Dup1
Dup2
Scenario 3 (*Fm_3)
~412 Kb
Dup1
Dup2
Dup1
Dup2


(a) Haplotype 1
Flank1
Dup1
Dup2
Flank1
Dup1 (~127Kb)
Int (~412 Kb)
Dup2 (~170Kb)
Flank2
T
T
T
T
T
T
G
G
TCCTGTCATCGGATAGCGCAGGCT
GACTCCAGATTCGGGTCGCGTAGGG
(b) Haplotype 2
Dup2
Dup1
Int
Flank1
Dup1 (~127Kb)
Int (~412 Kb)
Dup2 (~170Kb)
Flank2
A
A
A
A
G
G
G
G
C
C
C
C
AATCTCTGATAAGATATATGAATC
GACTCCAGATTCGGGTCGCGTAGGG


(a) Haplotype 1
Int
Dup2
Flank2
Flank1
Dup1 (~127Kb)
Int (~412 Kb)
Dup2 (~170Kb)
Flank2
A
A
A
A
A
A
AGGAGTGATCATAACCACTAGGAAA
(b) Haplotype 2
Dup1
Dup2
Dup1
Flank1
Dup1 (~127Kb)
Int (~412 Kb)
Dup2 (~170Kb)
Flank2
A
A
A
AATCTCTGATAAGATATATGAATC
G
G
G
T
T
TCCTGTCATCGGATAGCGCAGGCT
G
G
GACTCCAGATTCGGGTCGCGTAGGG


~500kb
~127kb
~412kb
~170kb
~500kb
Flank1
Dup1
Int
Dup2
Flank2
(a)
Kadaknath
(b)
Xichuan BBC
(c)
Yeonsan Ogye
(d)
Tianfu BBC
(e)
Ayam Cemani
(f)
Red Junglefowl
log (coverage)
5
6
7
8
9
10500000
11000000
11500000
12000000
Position along chromosome 20


(a)
Dup1
Dup2
BBC breeds
Kadaknath TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Yeonsan Ogye TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Xichuan BBC TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Ayam Cemani TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Tianfu BBC TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Non-BBC breeds
Red junglefowl TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCCTCCTTTCTGGCCCTCCCTAAGAGCGCTT CCCGCTGTCAGACCGTGGGGACAGCAGGGACTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Broiler TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCCTCCTTTCTGGCCCTCCCTAAGAGCGCTT CCCGCTGTCAGACCGTGGGGACAGCAGGGACTCACAGGGGCTGAGCAGCAGAGCTGAGCA
Gallus gallus murghi TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCCTCCTTTCTGGCCCTCCCTAAGAGCGCTT CCCGCTGTCAGACCGTGGGGACAGCAGGGACTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Tibetan fowl TACCCAGCGCTTTGTGGAAGTGGTTTTTTGCCTCCTTTCTGGCCCTCCCTAAGAGCGCTT CCCGCTGTCAGACCGTGGGGACAGCAGGGACTCACAGGGGCTGAGCAGCAGAGCTGAGCG
Flank1
Dup1
Int
Dup2
Flank2
(b)
Non-BBC breeds
Red junglefowl TCTGCTGTATCAGAAGAATGTTTCCAATTCTTGCCACTTCATATGTTTCCAATGAGAATT TCCTCCAGTAATATCTCCGAAGTTTTGCGGGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Broiler TCTGCTGTATCAGAAGAATGTTTCCAATTCTTGCCACTTCATATGTTTCCAATGAGAATT TCCTTCAGTAATATCTCCGAAGTTTTGCGGGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Gallus gallus murghi TCTGCTGTATCAGAAGAATGTTTCCAATTCTTGCCACTTCATATGTTTCCAATGAGAATT TCCTTCAGTAATATCTCCGAAGTTTTGCGGGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Tibetan fowl TCTGCTGTATCAGAAGAATGTTTCCAATTCTTGCCACTTCATATGTTTCCAATGAGAATT TCCTTCAGTAATATCTCCGAAGTTTTGCGGGTGTGAATTTGCCATCTCAGCTGTCTTTAC
BBC breeds
Kadaknath TCTGCTGTATCAGAAGAATGTTTCCAATTCGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Yeonsan Ogye TCTGCTGTATCAGAAGAATGTTTCCAATTCGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Xichuan BBC TCTGCTGTATCAGAAGAATGTTTCCAATTCGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Ayam Cemani TCTGCTGTATCAGAAGAATGTTTCCAATTCGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Tianfu BBC TCTGCTGTATCAGAAGAATGTTTCCAATTCGTGTGAATTTGCCATCTCAGCTGTCTTTAC
Dup1
Dup2


(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
R2
R1
Dup1
π
KADK
CHIN
0.006
0.004
0.002
0.000
iHS
4
3
2
1
0
-1
10000000
10500000
11000000
Position along chromosome 20
(c)
HM13
BPIL
TPX2
TTI1
DNTTIP1
ACOT8
SLC12A5