Common Ancestry of the *Id* Locus: Chromosomal Rearrangement and Polygenic Possibilities

Ashutosh Sharma, Nagarjun Vijay*

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

*Corresponding author: nagarjun@iiserb.ac.in

Abstract

The diversity in dermal pigmentation and plumage color among domestic chickens is striking, with Black Bone Chickens (BBC) particularly notable for their intense melanin hyperpigmentation. This unique trait is driven by a complex chromosomal rearrangement on chromosome 20 at the *Fm* locus, resulting in the overexpression of the *EDN3* (a gene central to melanocyte regulation). In contrast, the inhibition of dermal pigmentation is regulated by the *Id* locus. Although prior studies using genetic crosses, GWAS, and gene expression analysis have investigated the genetic underpinnings of the *Id* locus, its precise location and functional details remain elusive. Our study aims to precisely locate the *Id* locus, identify associated chromosomal rearrangements and candidate genes influencing dermal pigmentation, and examine the ancestral status of the *Id* locus in BBC breeds. Using public genomic data from BBC and non-BBC breeds, we refined the *Id* locus to a ~1.6 Mb region that co-localizes with Z amplicon repeat units at the distal end of the q-arm of chromosome Z within a 10.36 Mb inversion in Silkie BBC. Phylogenetic and population structure analyses reveal that the *Id* locus shares a common ancestry across all BBC breeds, much like the *Fm* locus. Selection signatures and highly differentiated BBC-specific SNPs within the *MTAP* gene position it as the prime candidate for the *Id* locus with *CCDC112* and additional genes, suggesting a possible polygenic nature. Our results suggest that the *Id* locus is shared among BBC breeds and may function as a supergene cluster in shank and dermal pigmentation variation.

Keywords: *Id* locus, Black-Bone Chickens, Dermal pigmentation, Chromosome Z, Genomic variants, Inversion

Introduction

Phenotypic diversity in pigmentation patterns within avian species has long been a focal point of biological research. The chicken has garnered considerable attention due to its economic and culinary significance and the remarkable variety of plumage colors and dermal pigmentation exhibited across different breeds. The most striking of such pigmentation phenotypes, fibromelanosis (*Fm*), refers to hyperpigmentation throughout the tissues and organs, encompassing the dermal layer of skin, bones, muscles, nerves, and gonads of black-bone chicken (BBC) (Dorshorst et al. 2011; Shinomiya et al. 2012). The underlying genetic cause of this hyperpigmentation, often referred to as the "black bone phenotype," was first investigated by Bateson and Punnett using the Silkie BBC breed (Bateson et al. 1911). Genetic crossing experiments identified two independent epistatically interacting factors influencing the degree of pigmentation observed in hybrid crosses between Silkie and White Leghorn chickens (Dunn and Jull 1927; Arora et al. 2011; Tian et al. 2014a). These two genetic factors responsible for dermal pigmentation are termed "fibromelanosis (*Fm*) locus" and "inhibitor of dermal melanin (*Id*) locus" (Mukherjee et al. 1969; Tixier-Boichard 2002).

The genetic basis of the *Fm* locus has been identified on chromosome 20 as a complex chromosomal rearrangement encompassing three different non-paralogous genomic regions, Dup1, Int, and Dup2, with the Dup1 and Dup2 regions being duplicated (Dorshorst et al. 2011; Shinomiya et al. 2012; Dharmayanthi et al. 2017; Sohn et al. 2018; Shinde et al. 2023). The most important region Dup1 contains five protein-coding genes: *EDN3*, *ZNF831*, *SLMO2*, *ATP5E*, and *TUBB1*. The prominent factor contributing to the intense melanization observed in the internal organs of BBC is the heightened expression of the Endothelin 3 gene (*EDN3*), which plays a pivotal role in producing melanocytes, specialized cells responsible for melanin synthesis (Baynash et al. 1994). The increased expression of *EDN3* resulting from the rearrangement is a key determinant in the hypermelanization phenomenon (Dorshorst et al. 2011).

BBC breeds are found globally, with a particularly strong presence in Asia. Comprehensive research, including genetic diversity studies, genome-wide analyses, and mitochondrial DNA

evidence, indicates that China is the primary origin of all BBC breeds (Zhang et al. 2018; Huang et al. 2021). Interestingly, BBC breeds from diverse geographic regions across Asia share a unique chromosomal rearrangement at the *Fm* locus (Shinde et al. 2023). A chromosome-level denovo genome assembly of a Silkie BBC individual proposed an alternative rearrangement (**Fm_1*) at the *Fm* locus (Zhu et al. 2023). However, a recent reanalysis of this data and additional data from other BBC breeds supports the presence of the common rearrangement (**Fm_2*) at the *Fm* locus (Sharma and Vijay 2024). The shared rearrangement and the population genetic signature of this region suggest a common ancestral origin for all BBC breeds.

In contrast to the autosomal dominant *Fm* locus, the mode of inheritance of *Id* is sex-linked incomplete dominance. As its name suggests, the *Id* locus is known to inhibit dermal pigmentation to different degrees based on its genetic state. Gene mapping studies have identified the physical location of the *Id* locus at the distal end of the q-arm of chromosome Z, in proximity to the sex-linked barring loci B (Bitgood 1985, 1988; Dorshorst and Ashwell 2009). The availability of genome assemblies helped identify that the chicken Z chromosome has a distinctive genetic structure known as the Z amplicon repeat unit (ZARU) at the distal end of the q-arm. The Z amplicon repeats consist of multiple copies of four genes: *ADCY10*, *C2orf3*, *ARHGAP33*, and *MRPL19* (Bellott et al. 2010; Yan et al. 2010) interspersed with tandem repeats (Edeen 1985).

Early draft genomes supported the localization of the *Id* locus to the distal end of the q-arm of the Z chromosome (Dorshorst et al. 2010). As genomic resources improved, the position of the *Id* locus was further refined. In recent years, utilizing a genome-wide SNP-trait association analysis, researchers pinpointed a potential *Id* locus in the galGal4 genome assembly, situated approximately at 78.4 Mb on chromosome Z (Dorshorst et al. 2010; Siwek et al. 2013; Li et al. 2014). Several studies have been undertaken involving crosses between BBC and non-BBC breeds, all in pursuit of identifying the candidate *Id* gene. These investigations have reported various potential genes which are associated with dermal and shank pigmentation variations (Mcgibbon 1978; Arora et al. 2011; Li et al. 2014; Tian et al. 2014b; Xu et al. 2017; Cha et al. 2023).

Numerous studies have been conducted to unravel the mysteries of the *Id* locus. Yet, the puzzle surrounding this locus persists, and the precise genomic region or gene responsible for its function remains elusive. However, with recent improvements in chicken genome assemblies and access to population genomic data from various BBC and non-BBC breeds, there is now an opportunity for a more comprehensive analysis to identify the *Id* locus accurately. Our study aims to address the following objectives: **1)** Assess the co-localization of Z amplicon repeats and *Id* locus along the Z chromosome. **2)** Evaluate the ancestral relationship at the *Id* locus among different BBC breeds. **3)** Identification of structural variation between BBC and non-BBC. **4)** Identify BBC-specific genomic signatures within the *Id* locus using population genomic statistics. **5)** Prioritize genomic variants associated with dermal pigmentation.

These objectives will guide a detailed investigation to advance our understanding of the *Id* locus and its role in pigmentation.

Materials and methods

Data collection

To comprehensively investigate the genomic structure and diversity of the *Id* locus, we acquired publicly accessible whole-genome resequencing short-read Illumina data of 270 chicken individuals. This dataset comprises three distinct wild junglefowl species: green, Ceylon, and grey. In addition to the wild junglefowl, the dataset includes multiple breeds of domesticated chicken. We included various BBC breeds, including Kadaknath, Silkie, Yeonsan Ogye, and Chinese breeds. Additionally, we incorporated non-black native chicken breeds like Aseel, Japanese gamefowl, Chinese gamefowl, Tibetan native chicken, Ethiopian native chicken, Sri Lankan native chicken, and Korean native chicken. The collection also encompasses commercial chicken breeds like White leghorn, Broiler, and Rhode Island Red.

Of the 270 individuals used in this study, 42 belonged to the BBC category, while 228 were categorized as non-BBC. We sourced these data from public repositories such as the National Center for Biotechnology Information (NCBI), the European Nucleotide Archive (ENA), the National Genomics Data Center (NGDC), and the Korean National Agricultural Biotechnology Information Center (KNABIC). The genome coverage of the resequencing data for each

individual was at least more than 10x. Comprehensive details for each individual can be found in **Table S1**.

Evaluation of assembly quality at the distal end of chromosome Z

To ensure the high quality of genome assemblies at the distal end of the q-arm of chromosome Z, we conducted assembly validation in the region of interest, specifically focusing on the potential *Id* locus. For this, we downloaded the high-coverage chicken ONT long-read data (SRR15421342, SRR15421343, SRR15421344, SRR15421345, SRR15421346, SRR13494713, and SRR13494714) of Huxu and Arbor Acres breeder rooster breed from ENA. We mapped these data to Huxu (GCA_024206055.2), GRCg6a (GCA_000002315.5) and GRCg7b (GCA_016699485.1) genome assemblies using the minimap2 v2.27-r1193 (Li, 2018) with --secondary=no --sam-hit-only --MD -Y -t 24 -ax map-ont flags.

To validate the continuity and integrity of these assemblies, we visually inspected a tiling path of long reads spanning the *Id* locus region using the UCSC genome browser (Karolchik et al. 2003; Fujita et al. 2011) for all three genome assemblies. To further evaluate the integrity of these genome assemblies, we used the Klumpy tool (Madrigal et al. 2024) on the aligned long reads using the “scan alignment” flag. We also compared all three genome assemblies, using SyRI (Synteny and Rearrangement Identifier) (Goel et al. 2019) to explore the structural similarities and rearrangements. To visualize the results obtained from SyRI, we utilized plotsr (plot structural rearrangements) (Goel and Schneeberger 2022). Our analysis found that the GRCg6a has better assembly quality (see **results** for details) compared to GRCg7b in the region of interest (*Id* locus). Therefore, we used GRCg6a as the reference genome assembly for subsequent studies.

Read mapping and variant calling

We mapped the paired-end raw Illumina reads from 270 individuals to the Gallus_gallus.GRCg6a genome assembly. The mapping was executed using the "bwa mem" flag, employing the default parameters within the BWA v 0.7.17-r1188 (Li and Durbin 2009). We incorporated read group information to enhance data integrity and eliminated duplicate reads from the resulting BAM files, utilizing the Picard tools (accessible at

https://github.com/broadinstitute/picard). We employed two distinct tools for variant calling: bcftools version 1.10.2 (Li and Durbin 2009; Danecek et al. 2021) and FreeBayes version 1.0.2 (Garrison and Marth 2012). For bcftools, we used various quality control flags such as -mapping quality -C 50, -min base quality -Q 20, and -min mapping quality -q 20. Similarly, for the FreeBayes, we imposed various quality control flags, including -min-alternate-count -C 10, -min-mapping-quality -m 20, -min-base-quality -q 20, and -min-coverage, requiring a minimum of 10 reads at a given locus. The aim was to identify SNPs with high quality and robustness.

To refine the dataset, we applied stringent filters using VCFtools version 0.1.16 (Danecek et al. 2011) to exclude low-quality variant calls and indels using the flags --remove-indels, --max-alleles 2, and --max-missing 0.95. Additionally, to enhance our SNP dataset's precision, we extracted common SNPs concordantly identified by both variant callers, minimizing the possibility of false positive SNP calls. After the filtration and obtaining the common SNPs, we ended with 62,102,065 SNPs genome-wide and 3,397,676 for chromosome Z.

Annotation of ZARU at the distal end of the q-arm of chromosome Z

In light of previous research (Bellott et al. 2010) about Z amplicon repeat units, we employed two distinct methodologies: (1) blastn and (2) read coverage estimation, to annotate the Z amplicon repeat units (ZARU) at the distal end of q-arm of chromosome Z in the GRCg6a genome assembly.

In the first approach, we utilized the blastn v2.9.0 (Altschul et al. 1990) to elucidate the spatial distribution of four specific genes: *ADCY10*, *C2orf3*, *ARHGAP33*, and *MRPL19*. To achieve this, we used the genomic sequences of all the annotated isoforms of these genes in the latest chicken genome as query sequences from NCBI for performing BLAST searches against the GRCg6a genome assembly. Subsequently, the obtained BLAST hits were saved in bed format, and overlapping hits were merged using the merge command in bedtools v2.27.1(Quinlan and Hall 2010). To visualize the distribution pattern of these genes, we divided chromosome Z into 1 Kb non-overlapping windows using bedtools. For each of these windows, we calculated the coverage of the BLAST hits using the bedtools coverage command.

In our second approach, we took a different perspective by evaluating read coverage in 10 Kb non-overlapping windows spanning chromosome Z. This allowed us to analyze read counts within each window, providing valuable confirmation regarding the presence of multiple copies of the four genes. Both methods were designed to yield robust evidence for annotating the ZARU regions. We visualized the distribution pattern and read coverage of ZARU regions using R v4.4.1 (Ihaka and Gentleman 1996), which allowed us to confirm and illustrate the ZARU layout across the distal end of the q-arm on chromosome Z.

Identification of chromosomal inversion in Silkie black-bone chicken

Following the annotation of ZARU at the distal end of chromosome Z, we examined structural variation between BBC and non-BBC breeds. Specifically, we examined the gapless CAU_Silkie and Huxu genome assemblies, focusing on potential structural rearrangements in this region. Pairwise alignment of these assemblies suggested the presence of an inversion at the distal end of chromosome Z.

We identified the genomic coordinates of each of the four ZARU genes on chromosome Z in both assemblies using blastn hits (see **Tables S2-S3**). Along with ZARU repeats, we also identified the tandem repeats across chromosome Z for both assemblies using the Global Repeat Map (GRM) application (Glunčić et al. 2022) (see **Table S4-S5**). Based on the identified ZARU gene repeats and tandem repeats, we constructed a map of repeat order for the distal end of chromosome Z.

Due to the presence of multiple copies of highly similar ZARU gene repeats, performing pairwise sequence alignment between the Huxu and Silkie assemblies in this region is challenging. To identify inversion boundaries, we compared the order of the repeat units, their lengths and orientations, and the relative distance between units of both assemblies. Through this manual comparison, we pinpointed rearranged junctions in Silkie BBC compared to the Huxu genome. We further validated this inversion in Silkie BBC by examining the junctions at each inversion boundary in detail. For Junction 1, we found a match spanning the initial segment of ZARU1 and the initial segment of inverted ZARU2; similarly, at Junction 2, we found a match spanning the inverted segment of ZARU1 with ZARU2. A comparison of repeat orders between

the Huxu and Silkie assemblies enabled us to identify inversion boundaries at a higher resolution.

To confirm that the identified matches were not due to random occurrence, we used a custom Python script (available in our GitHub repository). The script examines ZARU unit order, length, and distances between consecutive units in both strands to identify matches, allowing for both exact and tolerance-based matches to allow variation in unit lengths and distance between units to accommodate divergence between the breeds. Finally, the script calculates a p-value using a binomial test based on exact matches. The matched occurrences are then organized and output with details such as direction, number of exact and tolerance-based matches, p-value, and interval values, presented in a sorted format for clearer interpretation.

Phylogenetic, Principal component, and Admixture analysis

After validation of ZARU gene repeats distribution and inversion in Silkie BBC at the distal end of the q-arm of chromosome Z, we considered two single-copy regions, R1 (Z:78725000-80340000) and R2 (Z:81040000-82226500) which contains several protein-coding genes as potential *Id* locus regions. The R1 region is located entirely within the inverted region, while R2 has the same arrangement in BBC and non-BBC. The variant calls of the R1 and R2 regions was converted to phylip format using a custom Python script, vcf2phylip.py, downloaded from a GitHub repository (https://github.com/edgardomortiz/vcf2phylip). The resulting phylip file was then used as input for IQ-TREE 2 v2.3.4 (Minh et al. 2020) with -m MFP, --alrt 1000 -B 1000 --boot-trees flags to construct localized phylogenies for both regions. To visualize the generated phylogenetic trees, we employed Figtree v1.4.2 (Rambaut 2014). Notably, we observed the separation of BBC from non-BBC in R1, while no such separation was seen in the R2 region.

In the case of the *Fm* locus, the phylogenetic tree of Dup1 and Dup2 regions separated the BBC from non-BBC individuals. The structural rearrangement at the *Fm* locus reduces the gene flow between BBC and non-BBC breeds, which gives rise to this phylogenetic separation. Therefore, we reasoned that the *Id* locus may also have the BBC-specific phylogenetic signal due to a chromosomal rearrangement. Since the *Id* locus boundaries are unknown, we constructed 10Kb non-overlapping sliding window-based phylogenies for the entire chromosome Z using IQ-TREE 2 with the same settings. In total, we obtained 8252 window-based trees. We filtered out 1003

trees marked as "failed" by IQ-TREE's composition chi-square test due to significant deviations from average character composition, retaining only 7249 windows that passed the test. Visual interpretation from 7249 tree topologies is challenging. Therefore, we looked at the phylogenetic distance between these trees and the R1 topology. Using the 7249 qualified trees, we calculated the phylogenetic metrics distances between each of the 10Kb window trees and R1 region phylogeny using TreeCmp (Bogdanowicz et al. 2012). Specifically, five metrics (MatchingSplit, Robinson–Foulds, PathDifference, Quartet, and UMAST) suitable for unrooted phylogenies were utilized for the phylogenetic distance comparison. Finally, The results were visualized in R to explore the regions most closely associated with the R1 topology and are likely to be involved in the *Id* locus.

For the Principal Component Analysis (PCA) of both regions, we used PCAngsd v1.10 (Meisner and Albrechtsen 2018) based on genotype likelihood estimates derived from the Analysis of Next-Generation Sequencing Data (ANGSD) v0.935 (Korneliussen et al. 2014). The genotype likelihood values in beagle format, obtained from ANGSD, were employed in PCAngsd to determine principal components. Additionally, we estimated admixture proportions for both regions through NGSadmix (Skotte et al. 2013). In the NGSadmix analysis, we executed multiple runs with different values of K, ranging from K1 to K15, each with ten iterations. To discern the optimal K value, we applied the Evanno best K method (Evanno et al. 2005), an approach implemented in the Clumpak (Kopelman et al. 2015) web server, to identify the most appropriate K for our analysis. To visualize the results of PCA and Admixture analysis, we used R (Ihaka and Gentleman 1996).

SNP Density Analysis

To estimate the SNP density across the entire chromosome Z, We employed a 10Kb non-overlapping sliding window-based approach. The common SNPs identified by both variant callers were used as input in VCFtools with the --SNPdensity 10000 flag. Due to the poor quality of variant calls in the ZARU regions, which contain repetitive sequences, we identified callable sites using the CallableLoci flag available in GATK v3.8.1 (McKenna et al. 2010). The results of the SNP density analysis of all sites and callable sites were then visualized using R.

Genetic diversity and selection analysis

Our population genetic analysis exclusively included male (ZZ) individuals, a choice driven by the substantial genomic and sequence divergence observed in sex chromosomes. Using female (ZW) individuals can lead to uncertainty in the variant calls on the Z chromosome due to the mapping of gametologous or paralogous sequences shared between W and Z chromosomes (Ellegren 2011; Soh et al. 2014). Specifically, we selected three distinct breeds: two BBC breeds, Kadaknath (KADK) and Silkie (SILK), and a non-BBC breed represented by white leghorn chickens (WHLH). To ensure robust data quality, we carefully selected ten male individuals from each breed, each possessing sequencing coverage exceeding 20x (see **Table S6**). In pursuit of signatures of selection, we computed various population genetic statistics for the entire chromosome Z, employing 1 Kb windows as our analytical units. To assess genetic diversity within breeds, we utilized the folded site frequency spectrum (SFS) approach, implemented through ANGSD, to derive population-specific estimates of π (nucleotide diversity), Watterson's θ, and τ (Tajima's D). We calculated inter-population genomic differentiation (F_{ST}) using ANGSD to quantify genetic differentiation between populations. We calculated the weighted F_{ST} across chromosome Z by dividing the sum of the F_{ST} value of the BBC vs non-BBC comparisons by the BBC vs BBC F_{ST} value.

$$\frac{FST_{\text{KADK Vs. WHLH}} + FST_{\text{SILK Vs. WHLH}}}{FST_{\text{KADK Vs. SILK}}}$$

Additionally, we calculated divergence (D_{xy}) between population pairs using the all-sites variant files, employing the popgenWindows Python script, which can be found at https://github.com/simonhmartin/genomics_general. To identify regions potentially under selection, we defined $F_{ST,}$ weighted $F_{ST,}$ and D_{xy} outlier regions as 1 Kb windows falling within the top 10 percent of their respective estimates, signaling genomic regions of particular interest in population differentiation and evolutionary divergence.

Identification of micro-RNA, their target genes, and gene enrichment analysis

Based on our criteria of identifying genomic regions under selection using population genetic analysis, we found a 26 Kb long (Z:72580000-72606000) intergenic region in the Pre-ZARU region. We searched for micro-RNAs within this region using a nucleotide sequence-based search of micro-RNAs using the miRBase (Kozomara and Griffiths-Jones 2014) webserver. Our

search yielded the identification of five pre-micro-RNA precursors situated within the designated region. Subsequently, we sought to determine the target genes of these micro-RNAs. To accomplish this, we used the TargetScanHuman (Agarwal et al. 2015; McGeary et al. 2019) web server, specifying the chicken as the species of interest, and this enabled us to identify the target genes specific to these chicken-derived micro-RNAs. We performed gene enrichment analysis with the list of identified target genes. To explore the functional implications of these genes, we employed the KEGG pathway database (Kanehisa and Goto 2000) with default settings. This analysis was facilitated through the ShinyGO v 0.77 (Ge et al. 2020) web server, enabling us to gain insights into the potential pathways and biological processes influenced by these micro-RNAs and their associated target genes.

Identification of BBC-specific SNPs

To understand the impact of genetic variants within both coding and non-coding regions, we used the snpEff v4.3t along with GRCg6a.96 databases for annotation (Cingolani et al. 2012). This approach allowed us to comprehensively assess the genetic variants within regions putatively under selection in BBC and promoter regions of various genes. Our analysis was precisely honed to focus on alternative SNPs found exclusively in BBC individuals while their non-BBC counterparts retained reference alleles. We employed the bam-read count method to calculate the proportions of SNPs within all 270 individuals (42 BBC and 228 non-BBC) (Khanna et al. 2022). Additionally, based on a nucleotide sequence-based search in PROMO (Farré et al. 2003) (a virtual laboratory for the identification of putative transcription factor binding sites) using a 5 % maximum matrix dissimilarity rate, we identified the transcription factors which bind on the promoter region of *CCDC112* which harbours 3 SNPs highly differentiated in BBC breeds.

Results

Resolution of ZARU and its units is better in GRCg6a than in GRCg7b assembly

In the telomere-to-telomere Huxu genome assembly, the ZARU regions are fully assembled without gaps. This completeness of the Huxu genome is further supported by our independent ONT long-read alignment, as displayed in the UCSC browser screenshot (**Fig. S1**), and by the Klumpy scan alignment (**Fig. S2**). Using the Huxu assembly as a reference for the distal end of

the q-arm of chromosome Z, its comparison with GRCg6a and GRCg7b genome assemblies indicates that GRCg6a demonstrates higher assembly quality than GRCg7b (see **Fig. S3-S4**) in the distal end of chromosome Z. Specifically, ZARU2 is well-assembled in both the Huxu and GRCg6a assemblies, while in GRCg7b, ZARU2 appears to collapse with ZARU1 (see **Fig. S5-S6**). The Klumpy scan alignment further reveals that GRCg7b is more fragmented, containing more gaps relative to GRCg6a (see **Fig. S7-S8**). Although GRCg7b benefits from multi-platform, high-coverage data and exhibits generally high overall quality, its assembly of the ZARU regions remains suboptimal.

Id **locus is located between Z amplicon repeat units**

Based on the read coverage calculation of the entire chromosome Z, we observed high coverage in some regions located at the end of chromosome Z (**Fig. S9**). Using the BLASTN and read coverage approaches, we have refined the precise positions of Z-amplicon repeat units (located at Z:73613150-82529921) situated at the distal end of the q-arm of chromosome Z within the GRCg6a genome (**Fig. 1**) and GRCg7b assembly (**Fig. S10**). We have identified two single-copy regions, R1 and R2, which contain several protein-coding genes and divide the Z amplicon repeat units into three distinct segments (ZARU1, ZARU2 and ZARU3). R1, spanning from Z:78725000 to Z:80340000, extends over 1.61 Mb and harbors ten protein-coding genes (*MTAP*, *CDKN2A*, *CDKN2B*, *TRIM36*, *PGGT1B*, *CCDC112*, *FEMC1*, *ALDH7A1*, *GRAMD2B*, and *ZNF608)*. Similarly, R2, spanning from Z:81040000 to Z:82226500, encompasses a length of 1.18 Mb and hosts more than 15 protein-coding genes. Previous research (Li et al. 2014; Tian et al. 2014b; Xu et al. 2017; Cha et al. 2023), particularly concerning dermal pigmentation and plumage color variation, has highlighted several candidate genes predominantly located within the R1 region. Considering these earlier studies and our findings (discussed in the subsequent section), it becomes evident that the *Id* locus is co-localized with the Z-amplicon repeat units at the distal end of the q-arm of chromosome Z.

Id **locus co-localizes with an inversion within ZARU in Silkie BBC**

Through comparison of the organization of repeat units and tandem repeats on chromosome Z between the Huxu and Silkie assemblies, we observed that a substantial portion of ZARU1 (following breakpoint 1 at CP100594.1:75131357), along with the entire R1 region and the initial repeat units of ZARU2 (up to breakpoint 2 at CP100594.1:85496156), is inverted in Silkie

BBC (see **Fig. 2A**). In Silkie, the initial segment of the inverted ZARU2 joins with the initial part of ZARU1 at junction 1 (see **Fig. 2B**). Similarly, the segment of ZARU1 following breakpoint 1 connects with ZARU2 at junction 2 (**Fig. 2B**).

Along with the inversion, large regions of ZARU1 and ZARU2 were lost, resulting in a shorter ZARU region in Silkie BBC. A closer look at junction 1 reveals that the repeat order within ~74Kb in Silkie (a match of 59 units) occurs in two segments in Huxu (the initial segments of ZARU1 (a match of 41 units) and ZARU2 (a match of 34 units) located ~10.3Mb apart (see **Fig. 3** and **Table S7**). Likewise, the repeat order at junction 2, present within ~72Kb in Silkie (a match of 70 units), occurs in two parts in Huxu (ZARU1 (a match of 69 units) after breakpoint 1 and ZARU2 (a match of 26 units)), positioned ~10.6Mb apart (see **Fig. 4** and **Table S8**). These findings support the inversion of ZARU1 and the R1 region in Silkie BBC.

BBC breeds share common ancestry at *Id* locus

Compared to the genome-wide phylogeny (see **Fig. S11**), the local phylogeny of the rearranged *Fm* locus separates the BBC breeds from non-BBC (see **Fig. S12**) due to the reduced gene flow. As we identified that the entire R1 and ZARU1 have been inverted in BBC breeds, *Id* locus may also have the BBC-specific separation due to this inversion. Our higher-resolution identification of the *Id* locus is based on a comprehensive analysis of local phylogeny, principal component analysis (PCA), and admixture analysis of the R1 and R2 regions. Our findings suggest that R1 represents the *Id* locus for the following reasons: The genetic structure of the R1 region demarcates BBC from non-BBC breeds (**Fig. S13-S20**). In stark contrast, when examining the R2 region, we observe an intermingling of BBC and non-BBC breeds (**Fig. S21-S25**). This intermixing implies a lack of distinct differentiation between these two groups in the R2 region. The local phylogeny analysis of the R1, based on single nucleotide polymorphisms (SNP) data from 270 individuals, provides clear insights into the genetic relationships among different chicken breeds. As illustrated in **Fig. 5A** and **Fig. S13**, this analysis largely separates BBC from non-BBC individuals. Notably, a few individuals from the Tibetan native chicken population exhibit genetic proximity to Tibetan BBC individuals, suggesting potential gene flow. The phylogenetic distance of the 10Kb window trees within R1 to the R1 topology is much less than

the rest of the Z chromosome (see **Fig S14**). This pattern supports the conclusion that the entire R1 region possesses a distinct phylogenetic topology specific to BBC breeds.

Principal component analysis (PCA) of the R1 region further elucidates the genetic structure, with PC1 and PC2 accounting for 19.8% and 9.1% of the genetic variation, respectively. Notably, most BBC individuals form a distinct cluster, reinforcing the patterns observed in the local phylogeny analysis (**Fig. 5B** and **Fig. S15-17)**. An admixture analysis of the R1 region, aimed at elucidating ancestry information, reveals that BBC individuals are most distinct when the best value of K (genetic clusters) is set to 6. Non-BBC native breeds exhibit limited genetic overlap with BBC and commercial breeds, as depicted in **Fig. 5C** and **Fig. S18-S20**.

BBC-specific variants and genomic regions under selection

We employed two distinct approaches to find candidate genes or genomic regions associated with dermal pigmentation. In the first approach, we focused on identifying selective sweeps specific to BBC by leveraging various population genetic statistics. This analysis covered the distal end of the q-arm (Z:72000000-82529921). We sought to isolate the top 10% F_{ST} windows, derived from pairwise comparisons between BBC and non-BBC breeds, as potential regions subject to selection. To designate a region as being under selection, we established specific criteria. F_{ST} values were required to be less than 0.5 in comparisons involving BBC breeds, while in comparisons between BBC and non-BBC breeds, F_{ST} values should exceed 0.8. Furthermore, the target region must fall within the top 10% of weighted F_{ST} windows. This initial approach led us to identify numerous regions displaying substantial differentiation in BBC vs. non-BBC F_{ST} comparisons while at the same time having lower differentiation in BBC vs. BBC F_{ST} comparisons (**Fig. S26-28** and **Table S9** for detailed results).

In contrast to the signatures of selection, preexisting or standing genetic variation and breed-specific mutations arising from genetic drift can drive rapid evolution and influence phenotypic diversity and adaptation (Barrett and Schluter 2008). Consequently, our second approach investigates breed-specific genomic variants within BBC and non-BBC breeds. Our analysis has

revealed the presence of numerous genomic variants in both genic and intergenic regions that are nearly fixed in both BBC and non-BBC breeds (**Table S10**).

Putative BBC-specific selective sweep in the *MTAP* gene

Our genomic scan of R1 has unveiled distinct regions specific to BBC, characterized by a weighted F_{ST} exceeding the predefined threshold. As depicted in **Fig. S26** and **Table S9**, one particularly noteworthy region spans five Kb and is situated within the last intron of the *MTAP* gene. What sets this 5 Kb region apart is its markedly lower F_{ST} values in BBC vs. BBC comparisons compared to BBC vs. non-BBC comparisons, as evident in **Fig. 6A-D**. To further corroborate our findings, we examined D_{xy} results within the same region, and they reinforced the observed F_{ST} pattern, displaying striking differences, as depicted in **Fig. S29**. Within this 5 Kb region of the *MTAP* gene, we identified 6 BBC-specific SNPs, detailed in **Table S10**. Notably, the bam-read count data across 270 individuals underscore that these SNPs are nearly fixed among BBC individuals, irrespective of their geographic origins. A barplot depicting the percentage composition of each base at these six positions highlights the substantial difference in allele frequency between BBC and non-BBC individuals, as showcased in **Fig. 6E**.

BBC-specific variants in the promoter region of the *CCDC112* gene

Within the promoter region of the *CCDC112* gene, we identified three highly differentiated SNPs located at positions Z:79049203, Z:79049207, and Z:79049235. Our analysis across 42 BBC individuals found that ~83% are homozygous for the alternative allele, ~9% are heterozygous, and ~7% possess the reference allele. To provide a more accessible and insightful visualization of these findings, we've generated a barplot illustrating the bam-read count results for these three positions, as showcased in **Fig. 7**. In the case of the 228 non-BBC individuals, ~72% have reference alleles, ~12% are heterozygous, and ~ 14% possess alternative alleles, akin to those observed in BBC individuals. These large differences in allele frequency underscore the genetic divergence in this gene between BBC and non-BBC breeds. These three highly differentiated BBC-specific SNPs located within the promoter region of *CCDC112* overlap with binding sites of 27 transcription factors (see **Table S11**).

High differentiation of BBC-specific variants in R1 compared to the rest of chromosome Z

The ZARU regions on chromosome Z show a unique SNP distribution, likely influenced by poor short-read mapping quality, which can increase false SNP calls (see **Fig. S30**). Although SNP density in callable sites within ZARU regions is slightly lower than the rest of chromosome Z, the difference is minimal. Specifically, the R1 region exhibits BBC breed-specific variants with an SNP density of 0.043/base, similar to the overall chromosome Z density of 0.040/base.

In R1, we identified nine highly differentiated SNPs, six of which were in the *MTAP* gene and three in the *CCDC112* gene. This frequency of differentiated SNPs in R1 (0.0132%) notably exceeds the expected chrZ-wide proportion (0.00288%), where only 1.97 differentiated SNPs would be anticipated if distributed randomly. A binomial test confirmed the statistical significance of this excess, yielding a p-value of 0.00021, indicating that the observed concentration of differentiated SNPs in R1 is unlikely to be due to chance.

Pre-microRNA precursors identified in BBC-specific region target melanogenesis genes

We identified several other genomic regions in the Pre-ZARU region upstream with a selection signature in BBC breeds (**Table S9**). Among these, we found a 26 Kb continuous intergenic region spanning Z:72580000-72606000 with a remarkable pattern in pairwise F_{ST} and D_{xy} comparisons, as depicted in **Fig. 8A-D**. The examined D_{xy} results within the same region supported the observed F_{ST} pattern, displaying striking differences (see **Fig. S31**). The nucleotide diversity (π) within this region is higher in White Leghorn chickens. However, Kadaknath and Silkie chickens have markedly reduced nucleotide diversity, as visualized in **Fig. S32**.

Although this specific region lacks any protein-coding genes, we uncovered five pre-microRNA precursors highly similar to gga-mir-1457, gga-mir-6702, gga-mir-6551, gga-mir-6555, and gga-mir-6552 (**Fig. S33**). Of these five miRNAs, we found distinct BBC-specific SNPs within the paralogs of gga-mir-1457 and gga-mir-6552, which are predicted to target several genes (**Table S12-13**). Gene enrichment analysis of the targets of these two miRNAs revealed multiple target genes associated with various signaling pathways, including the GnRH signaling pathway, Apelin signaling pathway, Insulin signaling pathway, adrenergic signaling pathway, Wnt signaling pathway, and Hedgehog signaling pathway (**Fig. 8E-F,** and **Fig. S34-35**). Notably, our analysis in the context of gene enrichment unveiled an intriguing finding: 30 genes integral to the

melanogenesis pathway (see **Table S14)** serve as targets for these miRNAs, underscoring their potential role in regulating pigmentation-related processes.

Discussion

Our study has refined the position of the *Id* locus within a large (~10.36 Mb) inversion and identified several candidate *Id* locus genes. We demonstrate that the *Id* locus is co-localized among Z amplicon repeats, and the causative mutations likely occur within a region of ~1.61 Mb (i.e., R1) at the distal end of the q-arm of chromosome Z. The phylogenetic and population structure analysis of 270 individuals (42 BBC and 228 non-BBC) suggest that the *Id* locus has common ancestry among all BBC breeds. Within the refined region of the *Id* locus, we identified the BBC-specific regions putatively under selection. The two candidate genes, *MTAP* and *CCDC112*, involved in the melanogenesis pathway and plumage color variation, have BBC-specific SNPs in the intronic and promoter regions, respectively. The findings of our study suggest that the *Id* locus has a common ancestral state in BBC breeds and a polygenic nature, which is responsible for the variation in dermal pigmentation.

Chromosomal inversions are key in evolution, often suppressing recombination, limiting gene flow, and enabling the accumulation of breed or population-specific mutations, which contribute to developmental, morphological, and phenotypic diversity (Kirkpatrick and Barton 2006; Wellenreuther and Bernatchez 2018; Recuerda and Campagna 2024). Large inversions have been documented in birds, such as a 115 Mb inversion in quail (Sanchez-Donoso et al. 2022), four inversions ranging from 12-63 Mb in the zebra finch (Knief et al. 2016), and a 4.5 Mb-long inversion in the ruff (Küpper et al. 2015; Lamichhaney et al. 2015) linked with morphological variation, fitness and reproductive strategies. In chicken, a 7.4 Mb long inversion at chromosome 7 has been reported, which causes a rose-comb phenotype and poor sperm motility (Imsland et al. 2012; Wang et al. 2017). Overall, the role of chromosomal rearrangements in complex phenotypes is increasingly being recognised with the availability of genomic tools.

Our study has identified one of the longest novel chromosomal inversions (~10.36 Mb) at the highly repetitive distal end of chromosome Z in Silkie BBC. Our novel approach of using ZARU genes and tandem repeats as markers has shown that the *Id* locus is co-localized within Z amplicon repeat units and has undergone an inversion. This inversion could harbor genetic

changes causing inhibition of dermal pigmentation in BBC chickens. Identifying the causal mutation is challenging in such cases when a large inversion containing several genes is associated with a complex phenotype. Many sequence polymorphisms in strong linkage show a substantial association with the phenotype. Moreover, it is possible that multiple genes or multiple SNPs associated with the same gene could be responsible for the inhibition of dermal pigmentation. Since *Id* is sex-linked and has incomplete dominance, additional mutations contributing to the same inhibition phenotype could have accumulated within the inversion. Future studies relying on high-quality genome assemblies and comprehensive annotation of this region will be crucial for obtaining definitive answers.

The hyperpigmentation phenotype in the BBC breeds results from common complex chromosomal rearrangement at the *Fm* locus and distinguishes them from non-BBC breeds (Dharmayanthi et al. 2017; Shinde et al. 2023). The shared rearrangement in diverse BBC breeds supports a model of a common origin of BBC followed by dispersal globally. Therefore, we anticipated that the *Id* locus should possess unique characteristics specific to and shared among BBC breeds. Our focus on R1 as a potential *Id* locus led us to perform local phylogenetic analysis, principal component analysis (PCA), and admixture analysis. These analyses effectively separated the BBC breeds from the non-BBC breeds, emphasizing the common ancestry of the *Id* locus within the BBC breeds similar to the *Fm* locus. We observed relatedness at the *Id* locus between some BBC and Tibetan non-BBC native breeds, suggesting ongoing gene flow between these groups reported in earlier studies (Li et al. 2020). Hence, our analysis indicates that the *Id* locus emerged in BBC before their global dispersal and may share selection signatures and causative genetic variants involved in inhibiting dermal pigmentation.

One of our major findings is the signature of selection in the 5Kb genomic region of the last intron of *MTAP*, which contains 6 SNPs with an exceptionally high percentage of BBC-specific fixation. *MTAP* plays a crucial role in the melanogenesis pathway within melanocytes, the specialized cells responsible for producing the melanin pigment in the skin (Ainger et al. 2017). It is responsible for catalyzing the phosphorylation of methylthioadenosine (MTA), a byproduct of polyamine metabolism (Behrmann et al. 2003). MTA, in turn, hinders the function of phosphodiesterase (PDE), a regulator of intracellular cAMP levels. By blocking PDE, MTA blocks the cAMP-PKA-CREB pathway, leading to the inhibition of *MITF* stimulation. It has

been reported that mutations in the *MTAP* gene are associated with nevus count and melanoma development (Kvaskoff et al. 2011). Genomic variants identified in the *MTAP* are associated with plumage and dermal pigmentation in the F2 population crossed from White Leghorn and the Yeonsan Ogye (Heo et al. 2023; Cha et al. 2023). Functional gene expression analysis of Silkie and White Leghorn chicken embryos during early embryonic stages indicated distinct patterns of *MTAP* expression. Before embryonic day 10, White Leghorn embryos exhibited a higher *MTAP* expression level than Silkie embryos. However, a shift occurred after day 10, with Silkie embryos showing greater *MTAP* expression than White Leghorn embryos (Tian et al. 2014b). In the context of migration and differentiation studies of melanoblasts in Silkie and White Leghorn chickens, it was observed that Silkie neural crest cells underwent a slower melanocyte differentiation process compared to their White Leghorn counterparts. Specifically, on day 8, Silkie neural crest cells displayed less pigmentation than their White Leghorn counterparts. However, by day 11, they had achieved full pigmentation, signifying a distinct temporal pattern in melanocyte development between these two chicken breeds (Reedy et al. 1998; Faraco et al. 2001). The expression of *MTAP* and the migration and differentiation of melanocytes, both displaying a similar pattern, strongly imply the pivotal role of *MTAP* in regulating dermal pigmentation.

The sex-linked loci barring (B), dermal melanin inhibitor (*Id*), Translocational breakpoint (TB) and recessive white skin (y) are reported to occur in a linear order as B-*Id*-TB-y (Bitgood 1985, 1988). Subsequently, the sex-linked barring locus (B) was mapped to the *CDKN2A/2B* gene in the Barred Plymouth Rock chicken breed (Hellström et al. 2010). The *MTAP* gene occurs earlier (towards the centromere) than *CDKN2A/2B* in non-BBC breeds. Hence, the genomic position of the *MTAP* gene does not match with the B-*Id*-TB-y linear ordering of loci. However, after the inversion of the entire R1 region along with ZARU1, *CDKN2A/2B* is now located earlier than *MTAP* in BBC breeds. The rearrangement establishes the linear relationship of B-*Id*-TB-y, where *CDKN2A/2B* represents B, and *MTAP* can represent *Id*. Our findings of the R1 spanning inversion and putative selective sweep with SNPs suggest that *MTAP* is the most prominent candidate gene for the dermal melanin inhibitor.

SNPs in gene promoters can have implications for traits such as skin color and plumage in various species. For instance, in humans, an SNP in the promoter of the *SLC45A2* gene has been linked to variations in skin color (Graf et al. 2007). In ducks, the black plumage phenotype is determined by a mutation within the regulatory region of the *MC1R* gene (Liu et al. 2023). Similarly, studies on different chicken breeds have revealed that mutations in the promoter regions of genes like *MITF*, *TYR*, *ASIP*, and *GJA5* play a role in influencing skin color (Yu et al. 2017, 2019; Wang et al. 2018; Li et al. 2021). Interestingly, past research has established a connection between loss-of-function mutations in *CCDC112* and variations in feather color in geese (Ren et al. 2021). A previous gene expression analysis study conducted between Silkie BBC and White Leghorn chickens reported that after embryonic day 10, the *CCDC112* expression within the shank of Silkie embryos was greater than that of White Leghorn embryos (Tian et al. 2014b). This finding strongly implies a potential role of *CCDC112* in shank pigmentation in chickens. The regulatory mutations identified within *CCDC112's* promoter region hold promise for shedding light on their association with dermal pigmentation and plumage color. The transcription factors with binding sites overlapping three highly differentiated promoter SNPs in BBC breeds are primarily involved in developmental patterning (Lorenzo et al. 2017; Minto et al. 2024), cell differentiation (Darlington et al. 1998), immune response regulation (Tsukada et al. 2011; Villarino et al. 2015), and sex determination (Larney et al. 2014).

We found that paralogs of two miRNAs, gga-mir-1457 and gga-mir-6552, which have BBC-specific SNPs, target 30 genes associated with the melanogenesis pathway. Previous literature has extensively documented the role of miRNAs in melanocyte biology and pigmentation, primarily by modulating the expression of specific genes or pathways (Mione and Bosserhoff 2015). Associations between miRNAs and melanin production, skin pigmentation, and plumage coloration have been observed in diverse species such as Koi fishes (Tian et al. 2018), pigs (Yuan et al. 2023), alpacas (Zhu et al. 2019) and ducks (Apopo et al. 2015). Numerous miRNAs have been identified in the case of Muchuan BBC, impacting melanin production in the breast muscle (Yu et al. 2020). The miRNA in the pre-ZARU may be involved in plumage color regulation and require a comprehensive functional analysis, which can provide valuable insights into their roles within the melanogenesis pathway.

The Z Amplicon Repeat Units (ZARU) situated at the distal end of chromosome Z harbor hundreds of copies of protein-coding genes, including *ADCY10*, *C2orf3*, *ARHGAP33*, and *MRPL19*. Despite their abundance, the exact function of ZARU remains elusive, leaving open the possibility of additional structural rearrangements within this region that may be involved in complex epigenetic gene regulation and its potential association with the inhibition of dermal melanin production. For instance, the tandem repeats within the ZARU region are known to have an intricate methylation pattern (Eden et al. 1981). The repetitive nature of this region also poses significant challenges for mapping short reads and variant calling, which often leads to erroneous results (DePristo et al. 2011; Koboldt 2020). Consequently, analyzing the ZARU region under current technical conditions is challenging. Nonetheless, the availability of long-read data, advancements in mapping and variant calling pipelines hold promise for unraveling the mysteries surrounding this region in the future.

Conclusion

We explored the *Id* locus using read coverage analysis and blastn search for annotating amplicon repeats to map the organization of ZARU and localized population genetic structure analysis to refine its position to identify candidate genes. Our investigation concludes that the *Id* locus (Z:78725000-80340000) closely co-localizes with ZARU situated at the distal end of the q-arm of chromosome Z with a BBC-specific inversion. Multiple lines of evidence, such as local phylogeny, PCA, admixture and highly differentiated genomic variants, support the common ancestry of the R1 region within BBC breeds similar to the *Fm* locus. The population genetic signatures, the genome assembly of CAU_Silkie and the relative organisation of ZARU spanning the R1 region are consistent with a large chromosomal inversion in all BBC breeds. Among the identified genes, *MTAP* is the most prominent candidate for the *Id*, suggesting a pivotal role in inhibiting dermal pigmentation. In addition, *CCDC112* emerges as a potential candidate for influencing variations in pigmentation patterns. Further scrutiny of BBC breeds-

specific selective sweep region in the Pre-ZARU region led us to discover miRNAs that can target various genes involved in the melanogenesis pathway. Our findings collectively suggest that the *Id* locus is potentially polygenic and strongly associated with the inversion with shared ancestry among BBC breeds.

Availability of data

Scripts and data are available at https://github.com/Ashu2195/ID_locus.

CRediT authorship contribution statement

Ashutosh Sharma: Conceptualization, Formal analysis, Investigation, Visualization, Validation, Writing - original draft, Writing - review & editing. **Nagarjun Vijay:** Conceptualization, Resources, Writing - original draft, Writing - review & editing, Funding acquisition, Project administration, Supervision.

Acknowledgment

We thank 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.

Appendix

Supplementary 1. Supplementary figures

Supplementary 2. Supplementary tables

References

Agarwal V, Bell GW, Nam JW, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. Elife 4:. https://doi.org/10.7554/ELIFE.05005

Ainger SA, Jagirdar K, Katie J. Lee, et al (2017) Skin Pigmentation Genetics for the Clinic.

Dermatology

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Apopo S, Liu H, Jing L, et al (2015) Identification and profiling of microRNAs associated with white and black plumage pigmentation in the white and black feather bulbs of ducks by RNA sequencing. Anim Genet 46:627–635. https://doi.org/10.1111/AGE.12343

Arora G, Mishra SK, Nautiyal B, et al (2011) Genetics of hyperpigmentation associated with the Fibromelanosis gene (Fm) and analysis of growth and meat quality traits in crosses of native Indian Kadaknath chickens and non-indigenous breeds. Br Poult Sci 52:675–685. https://doi.org/10.1080/00071668.2011.635637

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. Trends Ecol Evol 23:38–44. https://doi.org/10.1016/J.TREE.2007.09.008

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

Baynash AG, Hosoda K, Giaid A, et al (1994) Interaction of endothelin-3 with endothelin-B receptor is essential for development of epidermal melanocytes and enteric neurons. Cell 79:1277–1285. https://doi.org/10.1016/0092-8674(94)90018-3

Behrmann I, Wallner S, Komyod W, et al (2003) Characterization of Methylthioadenosin Phosphorylase (MTAP) Expression in Malignant Melanoma. Am J Pathol 163:683–690. https://doi.org/10.1016/S0002-9440(10)63695-4

Bellott DW, Skaletsky H, Pyntikova T, et al (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 466:612–6. https://doi.org/10.1038/nature09172

Bitgood JJ (1985) Additional linkage relationships within the Z chromosome of the chicken. Poult Sci 64:2234–2238. https://doi.org/10.3382/PS.0642234

Bitgood JJ (1988) Linear relationship of the loci for barring, dermal melanin inhibitor, and

recessive white skin on the chicken Z chromosome. Poult Sci 67:530–533. https://doi.org/10.3382/ps.0670530

Bogdanowicz D, Giaro K, Wróbel B (2012) TreeCmp: Comparison of Trees in Polynomial Time. Evol Bioinform Online 8:475. https://doi.org/10.4137/EBO.S9657

Cha J, Jin D, Kim JH, et al (2023) Genome-wide association study revealed the genomic regions associated with skin pigmentation in an Ogye x White Leghorn F2 chicken population. Poult Sci 102:. https://doi.org/10.1016/J.PSJ.2023.102720

Cingolani P, Platts A, Wang LL, et al (2012) 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. https://doi.org/10.4161/FLY.19695

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/BIOINFORMATICS/BTR330

Danecek P, Bonfield JK, Liddle J, et al (2021) Twelve years of SAMtools and BCFtools. Gigascience 10:1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

Darlington GJ, Ross SE, MacDougald OA (1998) The role of C/EBP genes in adipocyte differentiation. J Biol Chem 273:30057–30060. https://doi.org/10.1074/jbc.273.46.30057

DePristo MA, Banks E, Poplin R, et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. https://doi.org/10.1038/ng.806

Dharmayanthi AB, Terai Y, Sulandari S, et al (2017) The origin and evolution of fibromelanosis in domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds. PLoS One 12:. https://doi.org/10.1371/JOURNAL.PONE.0173147

Dorshorst B, Molin AM, Rubin CJ, et al (2011) A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet 7:

Dorshorst B, Okimoto R, Ashwell C (2010) Genomic regions associated with dermal

hyperpigmentation, polydactyly and other morphological traits in the silkie chicken. J Hered

Dorshorst BJ, Ashwell CM (2009) Genetic mapping of the sex-linked barring gene in the chicken. Poult Sci 88:1811–1817. https://doi.org/10.3382/PS.2009-00134

Dunn LC, Jull MA (1927) On the inheritance of some characters op the silky fowl. J Genet. https://doi.org/10.1007/BF02983116

Edeen FC (1985) Truncated repeated sequences generated by recombination in a specific region. Biochemistry 24:229–233. https://doi.org/10.1021/BI00322A033

Eden FC, Musti AM, Sobieski DA (1981) Clusters of repeated sequences of chicken DNA are extensively methylated but contain specific undermethylated regions. J Mol Biol 148:129–151. https://doi.org/10.1016/0022-2836(81)90509-X

Ellegren H (2011) Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat Rev Genet 2011 123 12:157–166. https://doi.org/10.1038/nrg2948

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/J.1365-294X.2005.02553.X

Faraco CD, Soˆ S, Vaz SAS, et al (2001) Hyperpigmentation in the Silkie Fowl Correlates With Abnormal Migration of Fate-Restricted Melanoblasts and Loss of Environmental Barrier Molecules. Dev Dyn 220:212–22:. https://doi.org/10.1002/1097-0177

Farré D, Roset R, Huerta M, et al (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res 31:3651–3653. https://doi.org/10.1093/NAR/GKG605

Fujita PA, Rhead B, Zweig AS, et al (2011) The UCSC Genome Browser database: update 2011. Nucleic Acids Res 39:D876–D882. https://doi.org/10.1093/NAR/GKQ963

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

Ge SX, Jung D, Jung D, Yao R (2020) ShinyGO: A graphical gene-set enrichment tool for

animals and plants. Bioinformatics 36:2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Glunčić M, Vlahović I, Mršić L, Paar V (2022) Global Repeat Map (GRM) Application: Finding All DNA Tandem Repeat Units. Algorithms 15:458. https://doi.org/10.3390/A15120458/S1

Goel M, Schneeberger K (2022) plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics 38:2922–2926. https://doi.org/10.1093/BIOINFORMATICS/BTAC196

Goel M, Sun H, Jiao WB, Schneeberger K (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol 20:1–13. https://doi.org/10.1186/S13059-019-1911-0/FIGURES/5

Graf J, Voisey J, Hughes I, Van Daal A (2007) Promoter polymorphisms in the MATP (SLC45A2) gene are associated with normal human skin color variation. Hum Mutat 28:710–717. https://doi.org/10.1002/HUMU.20504

Hellström AR, Sundström E, Gunnarsson U, et al (2010) Sex-linked barring in chickens is controlled by the CDKN2A /B tumour suppressor locus. Pigment Cell Melanoma Res 23:521–530. https://doi.org/10.1111/J.1755-148X.2010.00700.X

Heo S, Cho S, Dinh PTN, et al (2023) A genome-wide association study for eumelanin pigmentation in chicken plumage using a computer vision approach. Anim Genet 54:355–362. https://doi.org/10.1111/AGE.13303

Huang X, Weng Z, He Y, et al (2021) Mitochondrial DNA diversity and demographic history of Black-boned chickens in China. Mitochondrial DNA Part B, Resour 6:1462–1467. https://doi.org/10.1080/23802359.2021.1912668

Ihaka R, Gentleman R (1996) R: A Language for Data Analysis and Graphics. J Comput Graph Stat 5:299–314. https://doi.org/10.1080/10618600.1996.10474713

Imsland F, Feng C, Boije H, et al (2012) The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. PLoS Genet

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27. https://doi.org/10.1093/NAR/28.1.27

Karolchik D, Baertsch R, Diekhans M, et al (2003) The UCSC Genome Browser Database. Nucleic Acids Res 31:51–54. https://doi.org/10.1093/NAR/GKG129

Khanna A, Larson DE, Srivatsan SN, et al (2022) Bam-readcount - rapid generation of basepair-resolution sequence metrics. ArXiv 7:3722. https://doi.org/10.21105/joss.03722

Kirkpatrick M, Barton N (2006) Chromosome Inversions, Local Adaptation and Speciation. Genetics 173:419–434. https://doi.org/10.1534/GENETICS.105.047985

Knief U, Hemmrich-Stanisak G, Wittig M, et al (2016) Fitness consequences of polymorphic inversions in the zebra finch genome. Genome Biol 17:1–22. https://doi.org/10.1186/S13059-016-1056-3/FIGURES/6

Koboldt DC (2020) Best practices for variant calling in clinical sequencing. Genome Med 2020 121 12:1–13. https://doi.org/10.1186/S13073-020-00791-W

Kopelman NM, Mayzel J, Jakobsson M, et al (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. https://doi.org/10.1111/1755-0998.12387

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:. https://doi.org/10.1186/S12859-014-0356-4

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:. https://doi.org/10.1093/NAR/GKT1181

Küpper C, Stocks M, Risse JE, et al (2015) A supergene determines highly divergent male reproductive morphs in the ruff. Nat Genet 2015 481 48:79–83. https://doi.org/10.1038/ng.3443

Kvaskoff M, Whiteman DC, Zhao ZZ, et al (2011) Polymorphisms in Nevus-Associated Genes MTAP, PLA2G6, and IRF4 and the Risk of Invasive Cutaneous Melanoma. Twin Res Hum Genet 14:422–432. https://doi.org/10.1375/TWIN.14.5.422

Lamichhaney S, Fan G, Widemo F, et al (2015) Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nat Genet 2015 481 48:84–88. https://doi.org/10.1038/ng.3430

Larney C, Bailey TL, Koopman P (2014) Switching on sex: transcriptional regulation of the testis-determining gene Sry. Development 141:2195–2205. https://doi.org/10.1242/DEV.107052

Li D, Sun G, Zhang M, et al (2020) Breeding history and candidate genes responsible for black skin of Xichuan black-bone chicken. BMC Genomics 21:. https://doi.org/10.1186/S12864-020-06900-8

Li G, Li D, Yang N, et al (2014) A genome-wide association study identifies novel single nucleotide polymorphisms associated with dermal shank pigmentation in chickens. Poult Sci 93:2983–2987. https://doi.org/10.3382/PS.2014-04164

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li J, Lee MO, Chen J, et al (2021) Cis-acting mutation affecting GJA5 transcription is underlying the Melanotic within-feather pigmentation pattern in chickens. Proc Natl Acad Sci U S A 118:e2109363118. https://doi.org/10.1073/PNAS.2109363118/SUPPL_FILE/PNAS.2109363118.SAPP.PDF

Liu H, Xi Y, Tang Q, et al (2023) Genetic fine-mapping reveals single nucleotide polymorphism mutations in the MC1R regulatory region associated with duck melanism. Mol Ecol 32:3076–3088. https://doi.org/10.1111/MEC.16924

Lorenzo PI, Juárez-Vicente F, Cobo-Vuilleumier N, et al (2017) The Diabetes-Linked Transcription Factor PAX4: From Gene to Functional Consequences. Genes 2017, Vol 8, Page 101 8:101. https://doi.org/10.3390/GENES8030101

Madrigal G, Minhas BF, Catchen J (2024) Klumpy: A tool to evaluate the integrity of long-read genome assemblies and illusive sequence motifs. Mol Ecol Resour e13982. https://doi.org/10.1111/1755-0998.13982

McGeary SE, Lin KS, Shi CY, et al (2019) The biochemical basis of microRNA targeting efficacy. Science (80-) 366:. https://doi.org/10.1126/SCIENCE.AAV1741

Mcgibbon WH (1978) The genetics of green spotting on shanks of Ancona chickens. J Hered 69:97–100. https://doi.org/10.1093/OXFORDJOURNALS.JHERED.A108912

McKenna A, Hanna M, Banks E, et al (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

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

Minh BQ, Schmidt HA, Chernomor O, et al (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/MOLBEV/MSAA015

Minto MS, Sotelo-Fonseca JE, Ramesh V, West AE (2024) Genome binding properties of Zic transcription factors underlie their changing functions during neuronal maturation. BMC Biol 22:1–20. https://doi.org/10.1186/S12915-024-01989-9/FIGURES/1

Mione M, Bosserhoff A (2015) MicroRNAs in melanocyte and melanoma biology. Pigment Cell Melanoma Res 28:340–354. https://doi.org/10.1111/PCMR.12346

Mukherjee TK, Jerome FN, Reinhart BS (1969) The Inhibition of Dermal Melanin in the Shanks of the Domestic Fowl. Poult Sci 48:2137–2142. https://doi.org/10.3382/PS.0482137

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842

Rambaut A (2014) FigTree v1.4.2., A Graphical Viewer of Phylogenetic Trees. In: http://tree.bio.ed.ac.uk/software/figtree/

Recuerda M, Campagna L (2024) How structural variants shape avian phenotypes: Lessons from model systems. Mol Ecol 33:e17364. https://doi.org/10.1111/MEC.17364

Reedy M V., Faraco CD, Erickson CA (1998) Specification and migration of melanoblasts at the vagal level and in hyperpigmented silkie chickens. Dev Dyn 213:476–485. https://doi.org/10.1002/(SICI)1097-0177(199812)213:4<476::AID-AJA12>3.0.CO;2-R

Ren S, Lyu G, Irwin DM, et al (2021) Pooled Sequencing Analysis of Geese (Anser cygnoides) Reveals Genomic Variations Associated With Feather Color. Front Genet 12:650013. https://doi.org/10.3389/FGENE.2021.650013/BIBTEX

Sanchez-Donoso I, Ravagni S, Rodríguez-Teijeiro JD, et al (2022) Massive genome inversion drives coexistence of divergent morphs in common quails. Curr Biol 32:462-469.e6. https://doi.org/10.1016/J.CUB.2021.11.019

Sharma A, Vijay N (2024) What is the correct genomic structure of the complex chromosomal rearrangement at the Fm locus in Silkie chicken? bioRxiv 2024.02.05.578760. https://doi.org/10.1101/2024.02.05.578760

Shinde SS, Sharma A, Vijay N (2023) Decoding the fibromelanosis locus complex chromosomal rearrangement of black-bone chicken: genetic differentiation, selective sweeps and protein-coding changes in Kadaknath chicken. Front Genet 14:1180658. https://doi.org/10.3389/FGENE.2023.1180658/BIBTEX

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

Siwek M, Wragg D, Sławińska A, et al (2013) Insights into the genetic history of Green-legged Partridgelike fowl: mtDNA and genome-wide SNP analysis. Anim Genet 44:522. https://doi.org/10.1111/AGE.12046

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

Soh YQS, Alföldi J, Pyntikova T, et al (2014) Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159:800–

813. https://doi.org/10.1016/J.CELL.2014.09.052

Sohn J Il, Nam K, Hong H, et al (2018) Whole genome and transcriptome maps of the entirely black native Korean chicken breed Yeonsan Ogye. Gigascience 7:1–14. https://doi.org/10.1093/GIGASCIENCE/GIY086

Tian M, Fang S, Wang Y, et al (2014a) Inverted duplication including Endothelin 3 closely related to dermal hyperpigmentation in Silkie chickens. Front Agric Sci Eng 1:121–129. https://doi.org/10.15302/J-FASE-2014026

Tian M, Hao R, Fang S, et al (2014b) Genomic regions associated with the sex-linked inhibitor of dermal melanin in Silkie chicken. Front Agric Sci Eng 1:242–249. https://doi.org/10.15302/J-FASE-2014018

Tian X, Pang X, Wang L, et al (2018) Dynamic regulation of mRNA and miRNA associated with the developmental stages of skin pigmentation in Japanese ornamental carp. Gene 666:32–43. https://doi.org/10.1016/J.GENE.2018.04.054

Tixier-Boichard M (2002) From phenotype to genotype: major genes in chickens. Worlds Poult Sci J 58:65–75. https://doi.org/10.1079/WPS20020008

Tsukada J, Yoshida Y, Kominato Y, Auron PE (2011) The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation. Cytokine 54:6–19. https://doi.org/10.1016/J.CYTO.2010.12.019

Villarino A V., Kanno Y, Ferdinand JR, O'Shea JJ (2015) Mechanisms of Jak/STAT Signaling in Immunity and Disease. J Immunol 194:21–27. https://doi.org/10.4049/JIMMUNOL.1401867

Wang G, Liao J, Tang M, Yu S (2018) Genetic variation in the MITF promoter affects skin colour and transcriptional activity in black-boned chickens. Br Poult Sci 59:21–27. https://doi.org/10.1080/00071668.2017.1379053

Wang Y, Li J, Feng C, et al (2017) Transcriptome analysis of comb and testis from Rose-comb Silky chicken (R1/R1) and Beijing Fatty wild type chicken (r/r). Poult Sci

Wellenreuther M, Bernatchez L (2018) Eco-Evolutionary Genomics of Chromosomal Inversions. Trends Ecol Evol 33:427–440. https://doi.org/10.1016/J.TREE.2018.04.002

Xu J, Lin S, Gao X, et al (2017) Mapping of Id locus for dermal shank melanin in a Chinese indigenous chicken breed. J Genet

Yan L, Liqing Z, Dexiang Z, et al (2010) Faster evolution of Z-linked duplicate genes in chicken. J Genet Genomics 37:695–702. https://doi.org/10.1016/S1673-8527(09)60087-4

Yu S, Liao J, Tang M, et al (2017) A functional single nucleotide polymorphism in the tyrosinase gene promoter affects skin color and transcription activity in the black-boned chicken. Poult Sci 96:4061–4067. https://doi.org/10.3382/PS/PEX217

Yu S, Wang G, Liao J (2019) Association of a novel SNP in the ASIP gene with skin color in black-bone chicken. Anim Genet 50:283–286. https://doi.org/10.1111/AGE.12768

Yu S, Wang G, Liao J, et al (2020) Identification of key microRNAs affecting melanogenesis of breast muscle in Muchuan black-boned chickens by RNA sequencing. Br Poult Sci 61:225–231. https://doi.org/10.1080/00071668.2019.1709619

Yuan W, Qin H, Bi H, et al (2023) Ssc-mir-221-3p regulates melanin production in Xiang pigs melanocytes by targeting the TYRP1 gene. BMC Genomics 24:1–14. https://doi.org/10.1186/S12864-023-09451-W/FIGURES/6

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

Zhu F, Yin Z-T, Zhao Q-S, et al (2023) A chromosome-level genome assembly for the Silkie chicken resolves complete sequences for key chicken metabolic, reproductive, and immunity genes. Commun Biol 2023 61 6:1–15. https://doi.org/10.1038/s42003-023-05619-y

Zhu Z, Ma Y, Li Y, et al (2019) Comparison of miRNA-101a-3p and miRNA-144a-3p regulation with the key genes of alpaca melanocyte pigmentation. BMC Mol Biol 20:1–11. https://doi.org/10.1186/S12867-019-0137-8/FIGURES/8

Figure legends:

Fig. 1. Magnified view of a ~10.5 Mb segment (72,000,000–82,529,921 in GRCg6a genome assembly) on the distal end of the q-arm of chicken chromosome Z: (A) Detailed view of three sets of Z amplicon repeat units (ZARU) within this segment, containing multiple copies of the genes *C2orf3*, *ARHGAP33*, and *MRPL19 in* ZARU1 and ZARU2. Multiple copies of The *ADCY10* gene are present at the start and the end of ZARU1, while various copies of *ARHGAP33* and *MRPL19* are present in ZARU3, as illustrated in the top panel. The second panel displays the short-read coverage across this region in 10 Kb non-overlapping windows, highlighting elevated coverage due to gene repeats. **(B)** Schematic of the ten genes located in the R1 region, with arrows indicating the transcriptional direction of each gene.

Fig. 2. Chromosomal rearrangement at the distal end of the q-arm of chromosome Z in Silkie (BBC) compared to Huxu (non-BBC): (A) Schematic representation of three Z amplicon repeat units (ZARU1, ZARU2, and ZARU3), along with the Pre-ZARU, R1, and R2 regions at the distal end of chromosome Z. ZARU1 and ZARU2 contain multiple copies of the C2orf3, ARHGAP33, and MRPL19 genes. In ZARU1, the ADCY10 gene has multiple copies at both the beginning and end of the unit, while additional copies of ARHGAP33 and MRPL19 are found within ZARU3. The red, dotted rectangular box and breakpoints 1 and 2 highlight the genomic region, and breakpoints 1 and 2 highlight the genomic region inverted in the Silkie BBC genome. **(B)** Diagram of the chromosomal rearrangement in Silkie BBC, where the majority of ZARU1 (8.806 Mb of 8.97 Mb), the entirety of the R1 region (~1.6 Mb), and the initial portion of ZARU2 (0.03 Mb) are inverted. This inversion results in a new configuration, with the initial portion of ZARU1 (0.191 Mb of 8.97 Mb) joined with the inverted segment of ZARU2 at Junction 1. The inverted ZARU1 is then joined with ZARU2 at Junction 2.

Fig. 3. Zoomed View of Junction 1 in Silkie Black Bone Chicken (BBC): (A) and (B) illustrate the initial portions of ZARU1 and the inverted ZARU2, respectively, in the Huxu non-BBC genome. (C) shows Junction 1 in the Silkie BBC, where the initial segment of ZARU1 and the inverted segment of ZARU2 have fused. Exons are indicated by vertical rectangular boxes, color-coded by gene: *ADCY10* (red), *MRPL19* (green), *C2orf3* (orange), and *ARHGAP33* (blue).

Tandem repeats are marked in dark cyan. The red vertical dotted line indicates the precise fusion point between the initial segment of ZARU1 and the inverted segment of ZARU2.

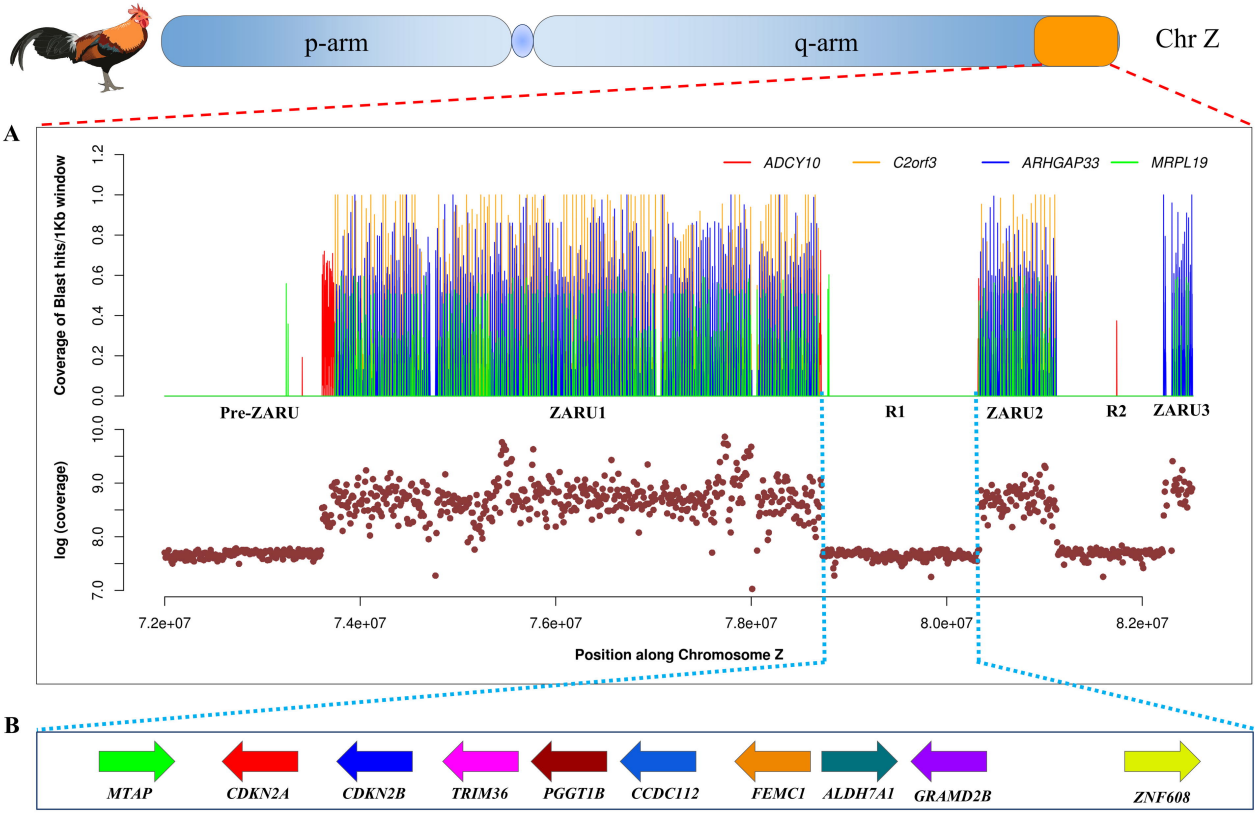
Fig. 4: Zoomed View of Junction 2 in Silkie Black Bone Chicken (BBC): (A) and (B) illustrate the inverted ZARU1 and ZARU2, respectively, in the Huxu non-BBC genome. (C) shows Junction 2 in the Silkie BBC, where the inverted ZARU1 and the ZARU2 have fused. Exons are indicated by vertical rectangular boxes, color-coded by gene: *ADCY10* (red), *MRPL19* (green), *C2orf3* (orange), and *ARHGAP33* (blue). Tandem repeats are marked in dark cyan. The red vertical dotted line indicates the precise fusion point between the inverted ZARU1 and the ZARU2.

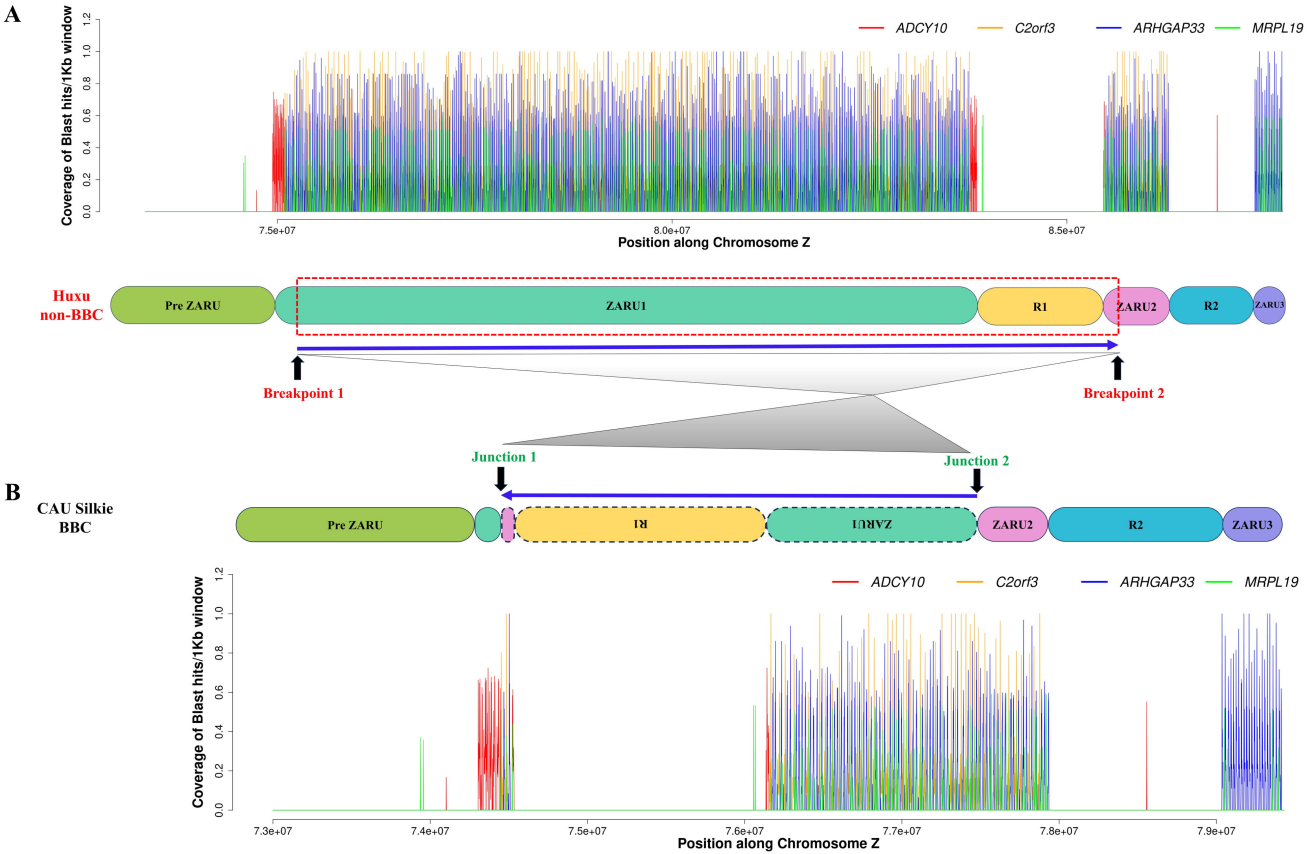
Fig. 5. Phylogenetic and Population Structure Analysis of the R1 Region: (A) Phylogenetic tree based on the R1 region, including 270 individuals (42 Black Bone Chickens and 228 non-black chickens), generated in iqtree2 and visualized in FigTree to illustrate genetic relationships among sampled populations. **(B)** Principal Component Analysis (PCA) of the R1 region, where the first principal component (PC1) captures 19.8% of the total genetic variance and the second principal component (PC2) accounts for 9.01%, highlighting genetic differentiation across populations. **(C)** Admixture analysis at K=6 reveals distinct ancestry proportions, with all Black Bone Chickens forming a unique genetic cluster, indicating significant genetic separation from non-black chickens.

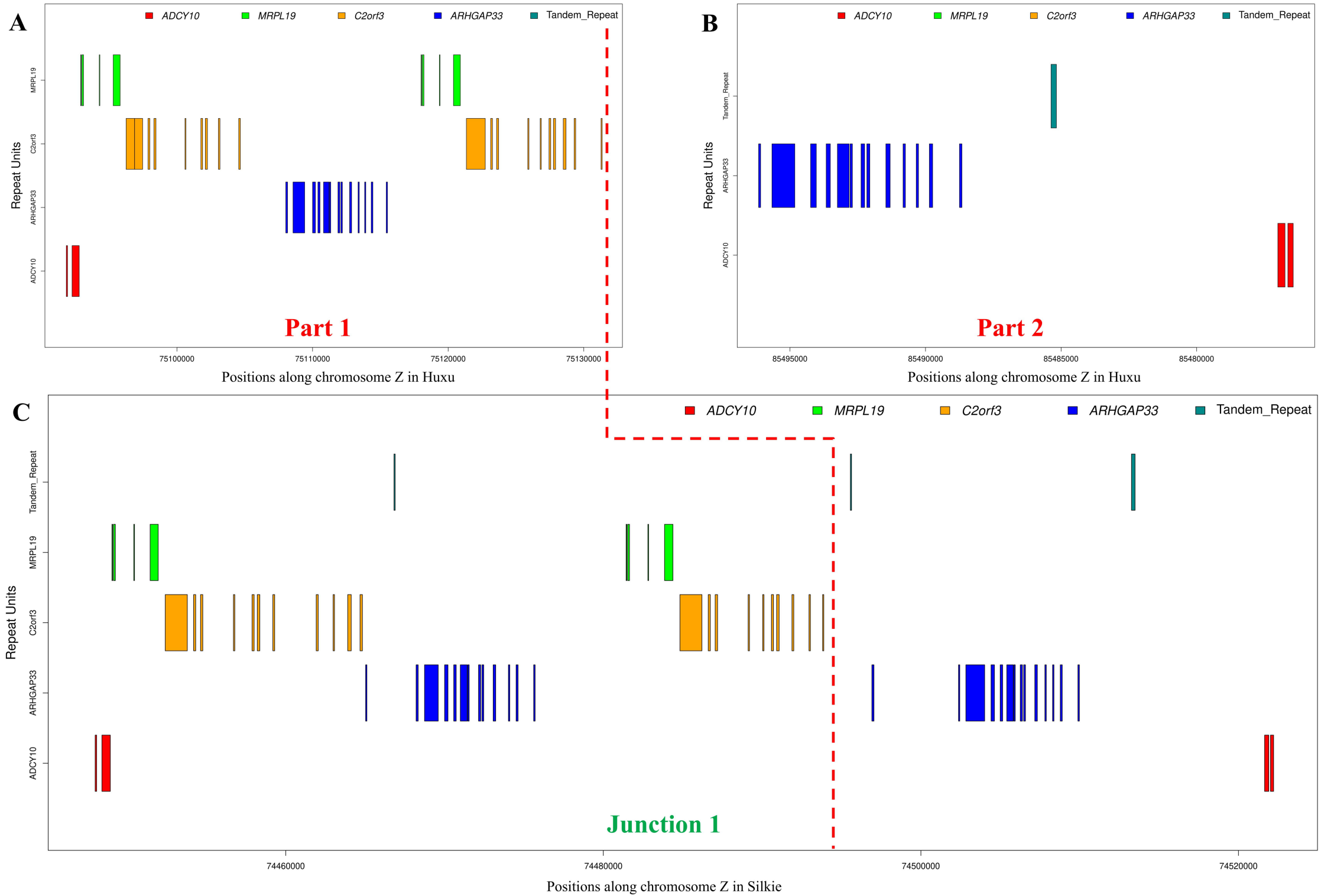
Fig. 6. Signature of selection in the last intron of the *MTAP* gene within the R1 region: Comparative analysis of weighted F_{ST} (calculated as the sum of BBC vs. non-BBC comparisons divided by BBC vs. BBC comparisons) and F_{ST} values, using 1Kb non-overlapping windows. A horizontal black dotted line represents the 90th percentile threshold for both Weighted F_{ST} and F_{ST} values, highlighting regions of high differentiation. **(A)** Weighted F_{ST} analysis comparing BBC (KADK and SILK) with non-BBC (WHLH) groups. **(B)** F_{ST} comparison between KADK and SILK breeds. **(C)** F_{ST} comparison between KADK and WHLH breeds. **(D)** F_{ST} comparison between SILK and WHLH breeds. **(E)** Six BBC-specific SNPs with high differentiation within a 5 kb region in the last intron of the *MTAP* gene were identified. The percentage of bases at each SNP position is shown for both BBC (n=42) and non-BBC (n=228) breeds, indicating breed-specific allele frequencies.

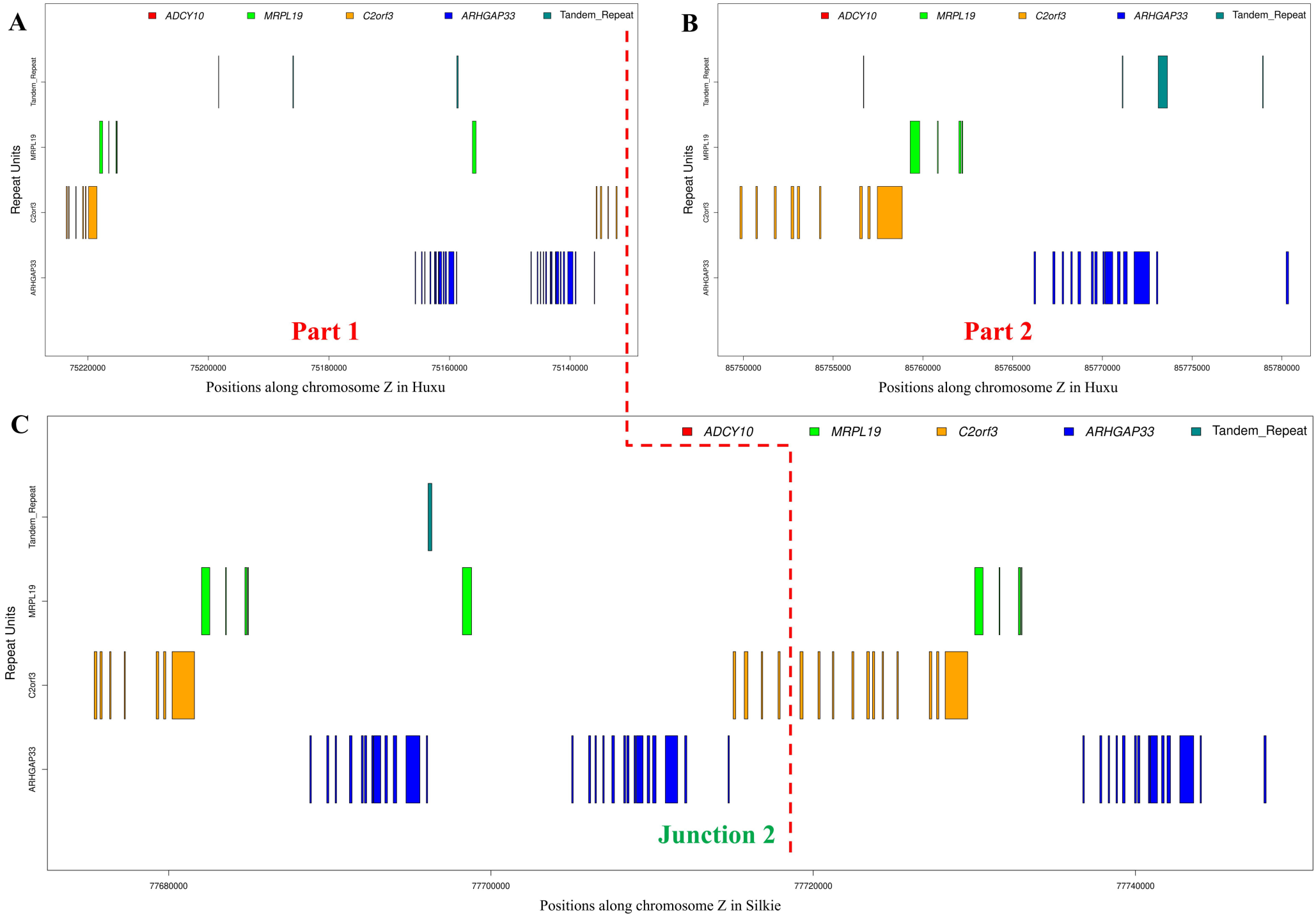
Figure 7. Black-bone chicken-specific SNPs in the promoter region of the *CCDC112* gene: (A) Illustration of three distinctive SNPs located within the promoter region of the *CCDC112* gene, unique to BBC, marked by orange triangles. **(B)** Allele frequency analysis shows the percentage of bases at each SNP position for the BBC breeds (n=42) and the non-BBC breeds (n=228), highlighting breed-specific genetic variation.

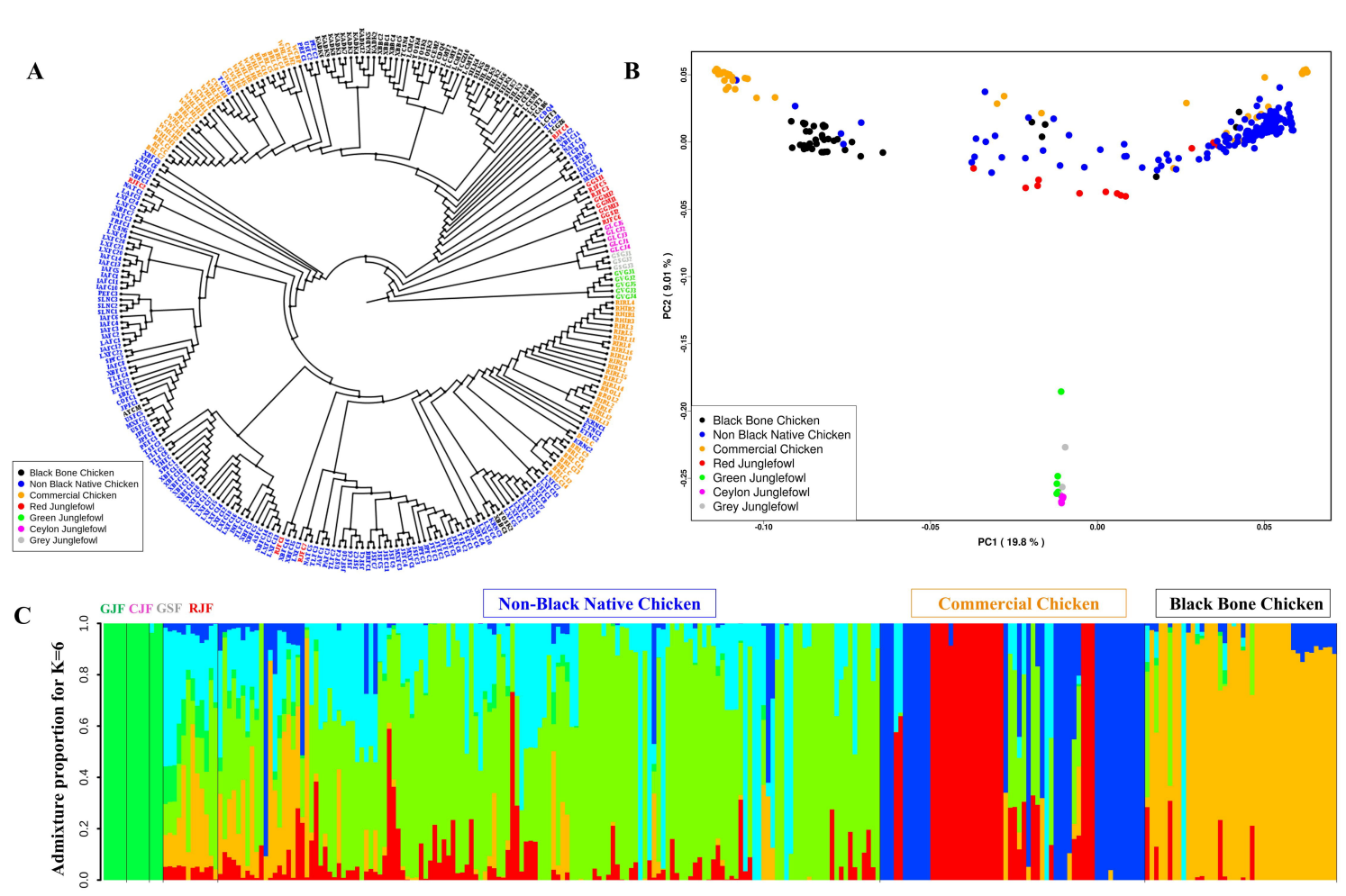
Fig. 8: Signature of selection in a 26 Kb long non-geneic region located in the Pre-ZARU region: Comparative analysis of weighted F_{ST} (calculated as the sum of BBC vs. non-BBC comparisons divided by BBC vs. BBC comparisons) and F_{ST} values, using 1Kb non-overlapping windows. A horizontal black dotted line represents the 90th percentile threshold for both Weighted F_{ST} and F_{ST} values, highlighting regions of high differentiation. **(A)** Weighted F_{ST} analysis comparing BBC (KADK and SILK) with non-BBC (WHLH) groups. **(B)** F_{ST} comparison between KADK and SILK breeds. **(C)** F_{ST} comparison between KADK and WHLH breeds. **(D)** F_{ST} comparison between SILK and WHLH breeds. **(E)** and **(F)** presents the gene enrichment analysis for the target genes of gga-mir-1457 and gga-mir-6552.

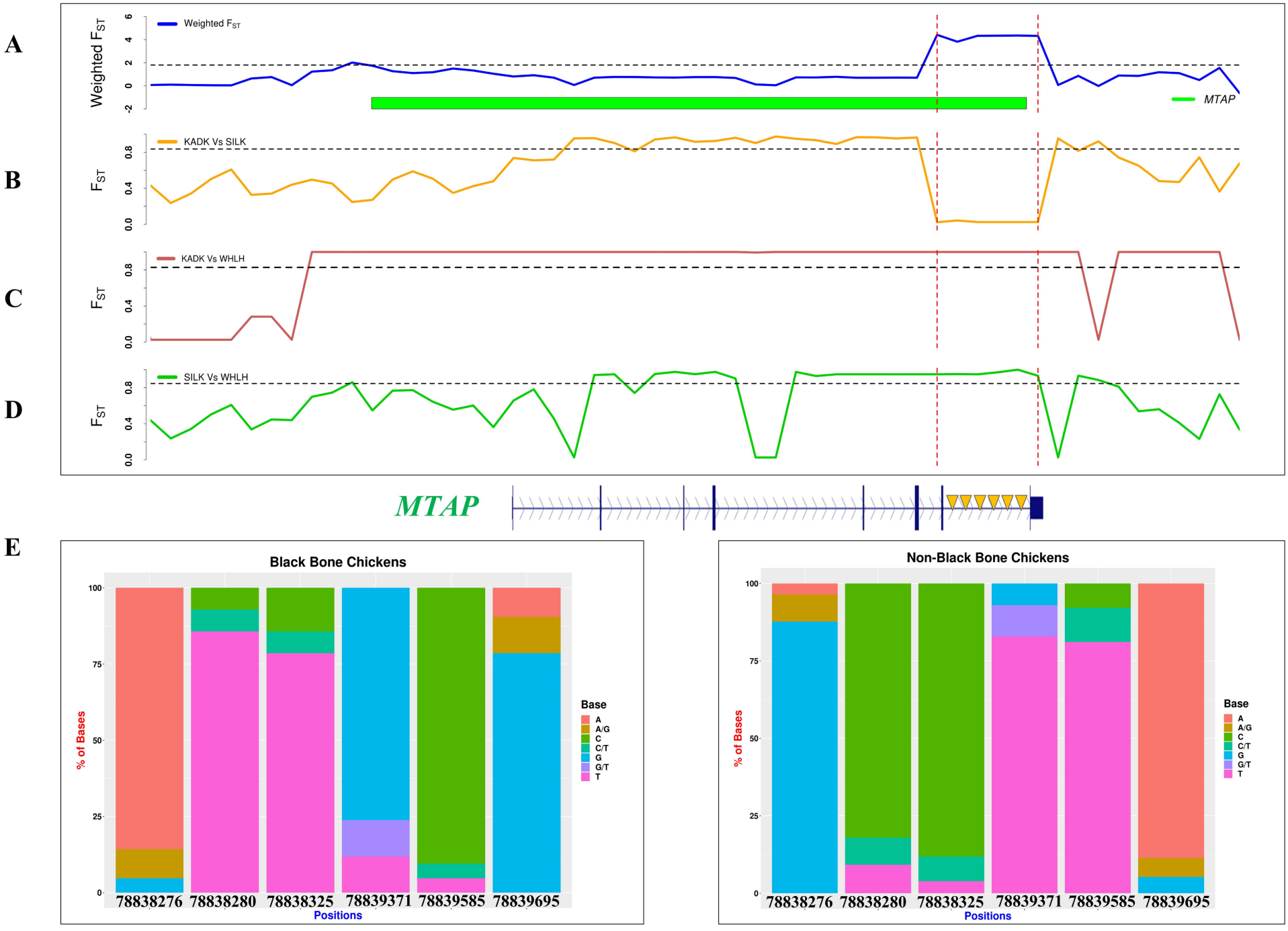

p-arm
q-arm
Chr Z
A
ADCY10
C2orf3
ARHGAP33
MRPL19
Coverage of Blast hits/1Kb window
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Pre-ZARU
ZARU1
R1
ZARU2
R2
ZARU3
log (coverage)
7.0
8.0
9.0
10.0
7.2e+07
7.4e+07
7.6e+07
7.8e+07
8.0e+07
8.2e+07
Position along Chromosome Z
B
MTAP
CDKN2A
CDKN2B
TRIM36
PGGT1B
CCDC112
FEMC1
ALDH7A1
GRAMD2B
ZNF608

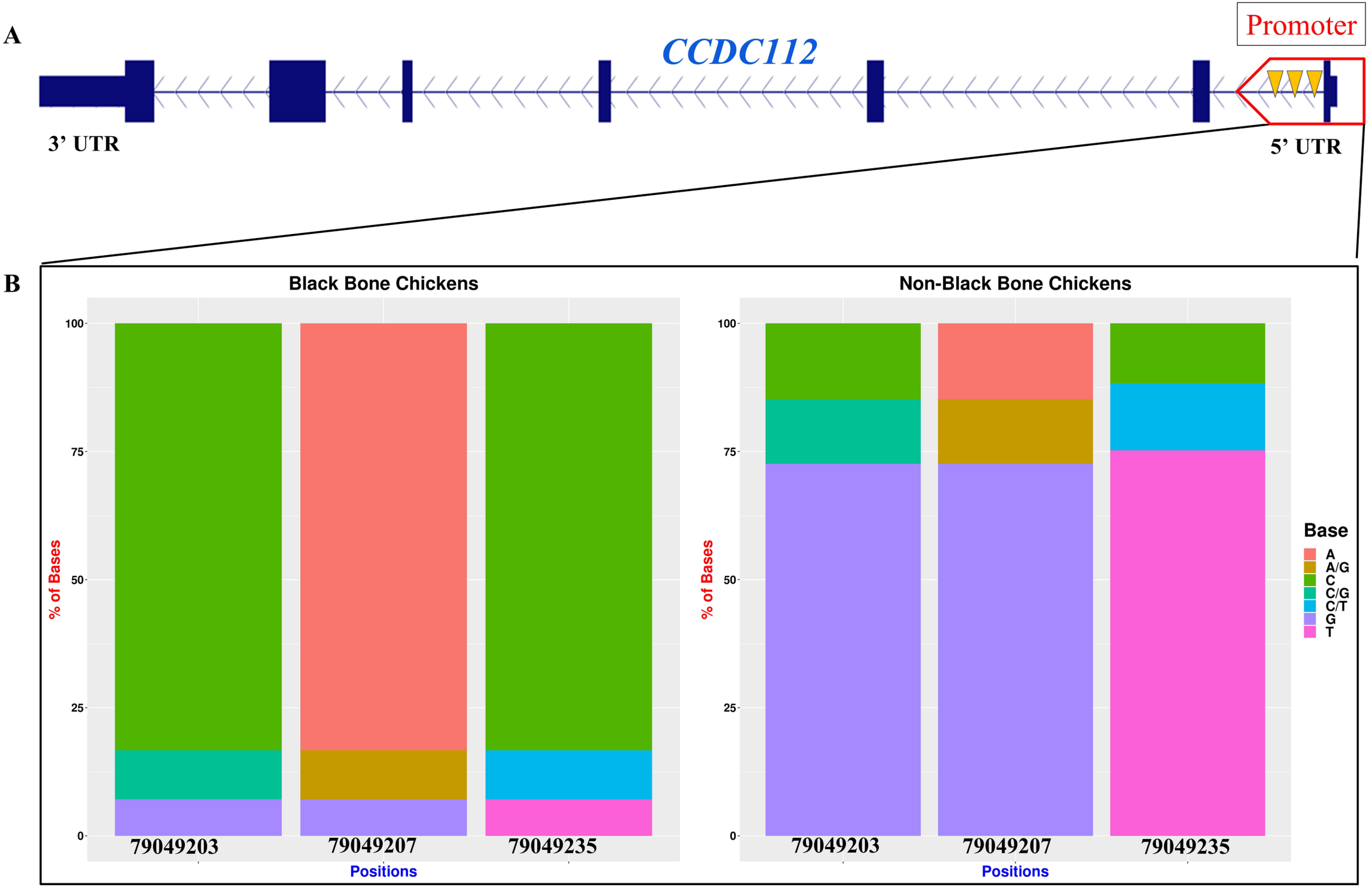

A
Coverage of Blast hits/1Kb window
0.0
0.2
0.4
0.6
0.8
1.0
1.2
ADCY10
C2orf3
ARHGAP33
MRPL19
7.5e+07
8.0e+07
8.5e+07
Position along Chromosome Z
Huxu
non-BBC
Pre ZARU
ZARU1
R1
ZARU2
R2
ZARU3
Breakpoint 1
Breakpoint 2
Junction 1
Junction 2
B
CAU Silkie
BBC
Pre ZARU
R1
ZARU1
ZARU2
R2
ZARU3
Coverage of Blast hits/1Kb window
0.0
0.2
0.4
0.6
0.8
1.0
1.2
ADCY10
C2orf3
ARHGAP33
MRPL19
7.3e+07
7.4e+07
7.5e+07
7.6e+07
7.7e+07
7.8e+07
7.9e+07
Position along Chromosome Z

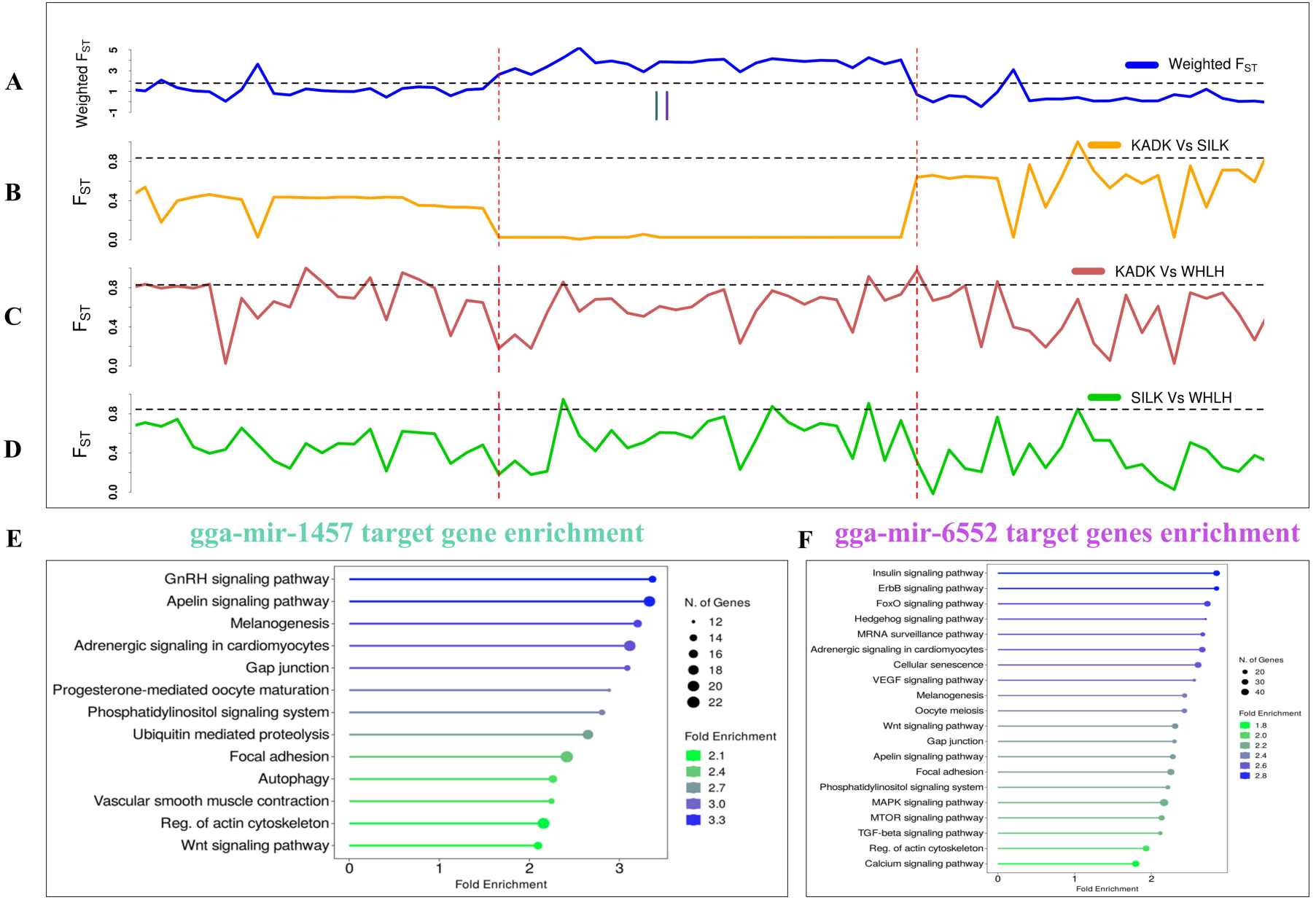

A
ADCY10
MRPL19
C2orf3
ARHGAP33
Tandem_Repeat
Repeat Units
Part 1
75100000
75110000
75120000
75130000
Positions along chromosome Z in Huxu
B
Part 2
85495000
85490000
85485000
85480000
Positions along chromosome Z in Huxu
C
Junction 1
74460000
74480000
74500000
74520000
Positions along chromosome Z in Silkie


A
ADCY10
MRPL19
C2orf3
ARHGAP33
Tandem_Repeat
Repeat Units
Tandem_Repeat
MRPL19
C2orf3
ARHGAP33
Part 1
75220000
75200000
75180000
75160000
75140000
Positions along chromosome Z in Huxu
B
ADCY10
MRPL19
C2orf3
ARHGAP33
Tandem_Repeat
Repeat Units
Tandem_Repeat
MRPL19
C2orf3
ARHGAP33
Part 2
85750000
85755000
85760000
85765000
85770000
85775000
85780000
Positions along chromosome Z in Huxu
C
ADCY10
MRPL19
C2orf3
ARHGAP33
Tandem_Repeat
Repeat Units
Tandem_Repeat
MRPL19
C2orf3
ARHGAP33
Junction 2
77680000
77700000
77720000
77740000
Positions along chromosome Z in Silkie


A
Black Bone Chicken
Non Black Native Chicken
Commercial Chicken
Red Junglefowl
Green Junglefowl
Ceylon Junglefowl
Grey Junglefowl
B
PC2 (9.01 %)
PC1 (19.8 %)
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-0.10
-0.05
0.00
0.05
Black Bone Chicken
Non Black Native Chicken
Commercial Chicken
Red Junglefowl
Green Junglefowl
Ceylon Junglefowl
Grey Junglefowl
C
GJF CJF GSF RJF
Non-Black Native Chicken
Commercial Chicken
Black Bone Chicken
Admixture proportion for K=6
1.0
0.8
0.6
0.4
0.2
0.0


A
Weighted F_{ST}
Weighted F_{ST}
MTAP
B
F_{ST}
KADK Vs SILK
C
F_{ST}
KADK Vs WHLH
D
F_{ST}
SILK Vs WHLH
MTAP
E
Black Bone Chickens
% of Bases
Positions
78838276
78838280
78838325
78839371
78839585
78839695
Base
A
A/G
C
C/T
G
G/T
T
Non-Black Bone Chickens
% of Bases
Positions
78838276
78838280
78838325
78839371
78839585
78839695
Base
A
A/G
C
C/T
G
G/T
T


A
CCDC112
Promoter
3' UTR
5' UTR
B
Black Bone Chickens
Non-Black Bone Chickens
% of Bases
100
75
50
25
0
79049203
79049207
79049235
Positions
Base
A
A/G
C
C/G
C/T
G
T


A
Weighted F_{ST}
Weighted F_{ST}
B
F_{ST}
KADK Vs SILK
C
F_{ST}
KADK Vs WHLH
D
F_{ST}
SILK Vs WHLH
E
gga-mir-1457 target gene enrichment
GnRH signaling pathway
Apelin signaling pathway
Melanogenesis
Adrenergic signaling in cardiomyocytes
Gap junction
Progesterone-mediated oocyte maturation
Phosphatidylinositol signaling system
Ubiquitin mediated proteolysis
Focal adhesion
Autophagy
Vascular smooth muscle contraction
Reg. of actin cytoskeleton
Wnt signaling pathway
Fold Enrichment
N. of Genes
Fold Enrichment
F
gga-mir-6552 target genes enrichment
Insulin signaling pathway
ErbB signaling pathway
FoxO signaling pathway
Hedgehog signaling pathway
MRNA surveillance pathway
Adrenergic signaling in cardiomyocytes
Cellular senescence
VEGF signaling pathway
Melanogenesis
Oocyte meiosis
Wnt signaling pathway
Gap junction
Apelin signaling pathway
Focal adhesion
Phosphatidylinositol signaling system
MAPK signaling pathway
MTOR signaling pathway
TGF-beta signaling pathway
Reg. of actin cytoskeleton
Calcium signaling pathway
Fold Enrichment
N. of Genes
Fold Enrichment