Common Ancestry of the *Id* Locus: Genetic Clues 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 of dermal pigmentation and plumage color in domestic chickens is intriguing. Black bone chickens (BBC) stand out with their intense melanin hyperpigmentation, driven by a complex chromosomal rearrangement on chromosome 20, specifically the *Fm* locus, causing overexpression of the *EDN3* gene. In contrast, dermal pigmentation inhibition is regulated by the sex-linked *Id* locus. Previous studies have used crosses, GWAS and gene expression analysis to identify the genetic basis of the *Id* locus, but the puzzle remains unsolved. Our study seeks to precisely locate the *Id* locus, identify gene(s) inhibiting dermal pigmentation, and explore its ancestral state in BBC breeds. We analyzed public data from 270 chickens (42 BBC and 228 non-BBC) to refine the position of the *Id* locus to a ~1.6 Mb region between Z amplicon repeats on the distal end of q-arm and found that it shares a common ancestry among all BBC breeds, much like the *Fm* locus. The selection signatures and high fixation levels within *MTAP*, *CCDC112* and a few other genes within this region support multiple candidates. Our results suggest that the *Id* locus is shared among BBC breeds and may have a polygenic nature, explaining the shank and dermal pigmentation variation.

Keywords: *Id* locus, Black-bone chickens, Dermal pigmentation, Chromosome Z, Polygenic

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 (Arora et al., 2011; Dunn and Jull, 1927; 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 (Dharmayanthi et al., 2017; Dorshorst et al., 2011; Shinde et al., 2023; Shinomiya et al., 2012; Sohn et al., 2018). 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). Notably, various BBC breeds from different geographic regions across Asia exhibit this distinctive chromosomal rearrangement at the *Fm* locus. The shared rearrangement and the population genetic signature of this region suggest a common ancestral origin for all BBC breeds (Shinde et al., 2023).

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, 1988, 1985; 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 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). Notably, parallel research on identifying the *Id* locus relying upon early draft genomes confirmed its localization at the distal end of the q-arm of chromosome Z (Dorshorst et al., 2010). Over time, with advancements in genomics and the availability of more comprehensive chicken genome data, the approximate position of the *Id* locus was 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; Li et al., 2014; Siwek et al.,

2013). 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 (Arora et al., 2011; Cha et al., 2023; Li et al., 2014; Mcgibbon, 1978; Tian et al., 2014b; Xu et al., 2017).

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. With recent advancements in the quality of the chicken genome and the availability of population genomic data from various BBC and non-BBC breeds, a newfound opportunity exists to conduct a comprehensive evaluation and accurate identification of the *Id* locus. 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)** Identify BBC-specific genomic signatures within the *Id* locus using population genomic statistics. **4)** Determine genomic variants associated with dermal pigmentation and plumage.

Results

Id locus is located between Z amplicon repeat units

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 Gallus_gallus.GRCg6a genome assembly (**Fig. 1**). We have identified two regions containing single-copy protein-coding genes, R1 and R2, within this region, which divide the Z amplicon repeat units into three distinct segments. R1, spanning from Z:78725000 to Z:80340000, extends over 1.61 Mb and harbors ten 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 (Cha et al., 2023; Li et al., 2014; Tian et al., 2014b; Xu et al., 2017), 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.

BBC breeds share common ancestry at *Id* locus

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. S1-10**). In stark contrast, when examining the R2 region, we observe an intermingling of BBC and non-BBC breeds (**Fig. S11-15**). 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. 2A** and **Fig. S1-4**, this analysis distinctly 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.

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, the BBC individuals form a distinct cluster, reinforcing the patterns observed in the local phylogeny analysis (**Fig. 2B** and **Fig. S5-7)**. 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. 2C** and **Fig. S8-10**.

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 selected 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 S16-18** and **Table S1** 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 S2**).

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. S16** and **Table S1**, 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. 3A-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. 3E-G**. Assessment of the evolutionary conservation of this region through the phastCons conservation score suggests a level of noteworthy conservation, as illustrated in **Fig. 3H**. Within this 5 Kb region of the *MTAP* gene, we identified 6 BBC-specific SNPs, detailed in **Table S2**. Notably, the bam-read count data across 270 individuals underscore that these SNPs are nearly ubiquitous among BBC individuals, irrespective of their geographic origins. A barplot depicting the percentage composition of each base at these six positions distinctly highlights the substantial difference in base composition between BBC and non-BBC individuals, as showcased in **Fig. 3I**.

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. 4**. In the case of the 228 non-BBC individuals, ~72% carry 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.

microRNAs in BBC-specific region target melanogenesis genes

We identified several other genomic regions upstream of the Z amplicon repeat units with a selection signature in BBC breeds (**Table S1**). 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. 5A-G**. The nucleotide diversity (π and Watterson's θ) within this region is higher in White Leghorn chickens. However, Kadaknath and Silkie chickens have markedly reduced nucleotide diversity, as visualized in **Fig. 5H-I**. An examination of the phastCons conservation score for this region indicated a notably high level of sequence conservation, as depicted in **Fig. 5J.**

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. S19**). 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 S3,4**). 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. 5K, L,** and **Fig. S20-21**). Notably, our analysis in the context of gene enrichment unveiled an intriguing finding: 30 genes integral to the melanogenesis pathway (see **Table S5)** 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 and identified several candidate genes responsible for inhibiting dermal melanin. We demonstrate that the *Id* locus is co-localized among Z amplicon repeats and occurs within a region of ~1.61 Mb at the distal end of the q-arm. 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.

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 breed-specific fixation. *MTAP* plays a crucial role in the melanogenesis pathway within melanocytes, the specialized cells responsible for producing the pigment melanin in the skin (Stephen A. Ainger et al., 2017). It is responsible for catalyzing the phosphorylation of methylthioadenosine (MTA), a byproduct of polyamine metabolism (Behrmann et al., 2003). Methylthioadenosine, in turn,

hinders the function of phosphodiesterase (PDE), a regulator of intracellular cAMP levels. By blocking PDE, Methylthioadenosine 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 (Cha et al., 2023; Heo 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 (Faraco et al., 2001; Reedy et al., 1998). 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. Our findings of 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 (Li et al., 2021; Wang et al., 2018; Yu et al., 2019, 2017). 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.

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). Out of the *Id* locus, these miRNAs may involved in plumage color and require a comprehensive functional analysis, which can provide valuable insights into their roles within the melanogenesis pathway.

The crosses between different BBC breeds and the gene expression analysis in early embryonic days can be used to study a polygenic basis for the *Id* locus. Hence, we propose a novel genetic crossing design for conducting a comprehensive study involving various BBC and non-BBC breeds (**Fig. 6**). The proposed experiment involves a round-robin cross between three BBC breeds (Kadaknath, Silkie, and Yeonsan Ogye) and two non-BBC breeds (White Leghorn and White Plymouth Rock) to generate the F1 crosses. The F1 individuals can be categorized based on dermal pigmentation and subsequently utilized for reciprocal crosses to generate F2 crosses. The F2 individuals can be further categorized into three groups based on their pigmentation levels, namely high, intermediate, and no pigmentation. Using GWAS, Population genomic study, and gene expression analysis during the early embryonic stages, particularly concerning genes containing morph-specific SNPs, may provide valuable insights into identifying the responsible genes. The approach we propose here involves multiple intermediate morphs

obtained from different genetic backgrounds (i.e., different BBC breeds) and can help identify shared genetic variants involved in dermal pigmentation.

Conclusion

We explored the *Id* locus and used read coverage, blastn search for amplicon repeats and localised population genetic structure to refine its position within the Gallus_gallus.GRCg6a genome assembly. Our investigation concludes that the *Id* locus (Z:78725000-80340000) closely co-localizes with the Z-amplicon repeat units situated at the distal end of the q-arm of chromosome Z. Multiple lines of evidence, such as local phylogeny, PCA, and admixture analysis, support this region's notion of common ancestry within BBC breeds like as *Fm* locus. Among the ten protein-coding genes in this region, *MTAP* stands out as 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 upstream to the Z amplicon led us to discover miRNAs which can target various genes involved in the melanogenesis pathway. Our findings collectively suggest that the *Id* locus is shared among BBC breeds, likely exhibiting a polygenic nature. Future studies considering the proposed novel genetic crossing design may reveal the mechanism of dermal inhibition at the molecular level.

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 S6**.

Read mapping and variant calling

We started our analysis by mapping 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 Burrows-Wheeler Aligner (BWA) software package version 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). For further analysis and variant calling, we subsetted the bam files for chromosome Z as our region of interest *Id* locus is known to be located on Z (Dorshorst et al., 2010). We employed two distinct tools for variant calling: bcftools version 1.10.2 (Danecek et al., 2021; Li and Durbin, 2009) and FreeBayes version 1.0.2 (Garrison and Marth, 2012). For bcftools, we used different 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, we used --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.

Validation of the Z amplicon repeat units 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 validate the presence of the Z amplicon repeat units (ZARU) at the distal end of q-arm of chromosome Z in the

Gallus_gallus.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 these genes as query sequences and performed BLAST searches against the Gallus_gallus.GRCg6a genome assembly. Subsequently, the obtained BLAST hits were saved in bed format. To visualize the distribution pattern of these genes, we divided chromosome Z into 1 Kb non-overlapping windows using bedtools v2.27.1 (Quinlan and Hall, 2010). 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 generating 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. This combined methodology aimed to deliver robust evidence and validate the Z amplicon repeat unit's existence.

Phylogenetic, Principal component, and Admixture analysis

After validation of ZARU at the distal end of the q-arm of chromosome Z in Gallus_gallus.GRCg6a genome assembly, we found two single-copy protein-coding regions between them. We considered these single-copy protein-coding regions R1 (Z:78725000-80340000) and R2 (Z:81040000-82226500) as potential *Id* locus regions. To construct a localized phylogeny of both regions, we used VCF2PopTree (Subramanian et al., 2019) and vk phylo implemented in the VCF-kit tool (Cook and Andersen, 2017). We applied Neighbour Joining (NJ) and the unweighted pair group method with arithmetic mean (UPGMA) to the SNPs identified within the R1 and R2 regions. Subsequently, for visualizing the resulting phylogenetic tree, we utilized Figtree v1.4.2 (Rambaut, 2014).

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).

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 S7**). 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, dividing the sum of the F_{ST} value of BBC vs. BBC comparison's F_{ST} value by the non-BBC comparison.

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

Additionally, we calculated divergence (D_{xy}) between population pairs, 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 upstream of the ZARU. 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).

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.

Conflict of interest statement

The authors declare that they have no conflict of interest.

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, G.W., Nam, J.W., Bartel, D.P., 2015. Predicting effective microRNA target sites in mammalian mRNAs. Elife 4. https://doi.org/10.7554/ELIFE.05005

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 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., Du, X., Xie, S., Gong, Y., Xu, R., Li, S., 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, S.K., Nautiyal, B., Pratap, S.O., Gupta, A., Beura, C.K., Singh, D.P., 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, R.D.H., 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, B.W., -a, F., Punnett, R.C., 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, A.G., Hosoda, K., Giaid, A., Richardson, J.A., Emoto, N., Hammer, R.E., Yanagisawa, M., 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., Heinrich, P.C., Schuierer, M., Buettner, R., Bosserhoff, A.K., 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, D.W., Skaletsky, H., Pyntikova, T., Mardis, E.R., Graves, T., Kremitzki, C., Brown, L.G., Rozen, S., Warren, W.C., Wilson, R.K., Page, D.C., 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, J.J., 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

Bitgood, J.J., 1985. Additional linkage relationships within the Z chromosome of the chicken.

Poult. Sci. 64, 2234–2238. https://doi.org/10.3382/PS.0642234

Cha, J., Jin, D., Kim, J.H., Kim, S.C., Lim, J.A., Chai, H.H., Jung, S. a., Lee, J.H., Lee, S.H., 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, 102720. https://doi.org/10.1016/J.PSJ.2023.102720

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 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

Cook, D.E., Andersen, E.C., 2017. VCF-kit: assorted utilities for the variant call format. Bioinformatics 33, 1581–1582. https://doi.org/10.1093/BIOINFORMATICS/BTX011

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/BIOINFORMATICS/BTR330

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10, 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

Dharmayanthi, A.B., Terai, Y., Sulandari, S., Zein, M.S.A., Akiyama, T., Satta, Y., 2017. The origin and evolution of fibromelanosis in domesticated chickens: Genomic comparison of Indonesian Cemani and Chinese Silkie breeds. PLoS One 12, e0173147. https://doi.org/10.1371/JOURNAL.PONE.0173147

Dorshorst, B., Molin, A.M., Rubin, C.J., Johansson, A.M., Strömstedt, L., Pham, M.H., Chen, C.F., Hallböök, F., Ashwell, C., Andersson, L., 2011. A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 7. https://doi.org/10.1371/journal.pgen.1002412

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

hyperpigmentation, polydactyly and other morphological traits in the silkie chicken. J. Hered. 101, 339–350. https://doi.org/10.1093/jhered/esp120

Dorshorst, B.J., Ashwell, C.M., 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, L.C., Jull, M.A., 1927. On the inheritance of some characters op the silky fowl. J. Genet. 19, 27–63. https://doi.org/10.1007/BF02983116/METRICS

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, C.D., Soˆ, S., Vaz, S.A.S., Vé, M., Pa´stor, R.D., Erickson, C.A., 2001. Hyperpigmentation in the Silkie Fowl Correlates With Abnormal Migration of Fate-Restricted Melanoblasts and Loss of Environmental Barrier Molecules. Dev. Dyn. 220:212–225. https://doi.org/10.1002/1097-0177

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

Ge, S.X., 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

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

Heo, S., Cho, S., Dinh, P.T.N., Park, J., Jin, D.H., Cha, J., Kim, Y.K., Koh, Y.J., Lee, S.H., Lee, J.H., 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

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

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

Khanna, A., Larson, D.E., Srivatsan, S.N., Mosior, M., Abbott, T.E., Kiwala, S., Ley, T.J., Duncavage, E.J., Walter, M.J., Walker, J.R., Griffith, O.L., Griffith, M., Miller, C.A., 2022. Bam-readcount - rapid generation of basepair-resolution sequence metrics. ArXiv 7, 3722. https://doi.org/10.21105/joss.03722

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 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, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 1–13.

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

Kvaskoff, M., Whiteman, D.C., Zhao, Z.Z., Montgomery, G.W., Martin, N.G., Hayward, N.K., Duffy, D.L., 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

Li, D., Sun, G., Zhang, M., Cao, Y., Zhang, C., Fu, Y., Li, F., Li, G., Jiang, R., Han, R., Li, Z., Wang, Y., Tian, Y., Liu, X., Li, W., Kang, X., 2020. Breeding history and candidate genes responsible for black skin of Xichuan black-bone chicken. BMC Genomics 21, 1–15. https://doi.org/10.1186/S12864-020-06900-8/FIGURES/8

Li, G., Li, D., Yang, N., Qu, L., Hou, Z., Zheng, J., Xu, G., Chen, S., 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, M.O., Chen, J., Davis, B.W., Dorshorst, B.J., Siegel, P.B., Inaba, M., Jiang, T.X., Chuong, C.M., Andersson, L., 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., Qi, J., Zhou, Z., Guo, Z., Fan, W., Hu, J., Xu, Y., Liang, S., Xie, M., Tang, J., Zhang, Y., Hou, S., 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

McGeary, S.E., Lin, K.S., Shi, C.Y., Pham, T.M., Bisaria, N., Kelley, G.M., Bartel, D.P., 2019. The biochemical basis of microRNA targeting efficacy. Science (80-.). 366. https://doi.org/10.1126/SCIENCE.AAV1741

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

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

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, T.K., Jerome, F.N., Reinhart, B.S., 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, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033

Rambaut, A., 2014. FigTree v1.4.2., A Graphical Viewer of Phylogenetic Trees. [WWW Document]. http://tree.bio.ed.ac.uk/software/figtree/. URL http://tree.bio.ed.ac.uk/software/figtree/

Reedy, M. V., Faraco, C.D., Erickson, C.A., 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, D.M., Liu, X., Feng, C., Luo, R., Zhang, J., Sun, Y., Shang, S., Zhang, S., Wang, Z., 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

Shinde, S.S., 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., Mizutani, M., Namikawa, T., Matsuda, Y., Akiyama, T., 2012. Gene Duplication of endothelin 3 Is Closely Correlated with the Hyperpigmentation of the Internal Organs (Fibromelanosis) in Silky Chickens. Genetics 190, 627–638. https://doi.org/10.1534/GENETICS.111.136705

Siwek, M., Wragg, D., Sławińska, A., Malek, M., Hanotte, O., Mwacharo, J.M., 2013. Insights into the genetic history of Green-legged Partridgelike fowl: mtDNA and genome-wide SNP analysis. Anim. Genet. 44, 522–532. https://doi.org/10.1111/AGE.12046

Skotte, L., Korneliussen, T.S., 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, Y.Q.S., Alföldi, J., Pyntikova, T., Brown, L.G., Graves, T., Minx, P.J., Fulton, R.S., Kremitzki, C., Koutseva, N., Mueller, J.L., Rozen, S., Hughes, J.F., Owens, E., Womack, J.E., Murphy, W.J., Cao, Q., De Jong, P., Warren, W.C., Wilson, R.K., Skaletsky, H., Page, D.C., 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., Kim, J.M., Lim, D., Lee, K.T., Do, Y.J., Cho, C.Y., Kim, N., Chai, H.H., Nam, J.W., 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

Stephen A. Ainger, Kasturee Jagirdar, Katie J. Lee, H. Peter Soyer, Richard A. Sturm, 2017. Skin Pigmentation Genetics for the Clinic. Dermatology.

Subramanian, S., Ramasamy, U., Chen, D., 2019. VCF2PopTree: a client-side software to construct population phylogeny from genome-wide SNPs. PeerJ 7. https://doi.org/10.7717/PEERJ.8213

Tian, M., Fang, S., Wang, Y., Gu, X., Feng, C., Hao, R., Hu, X., Li, N., 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., Wang, Y., Gu, X., Feng, C., Hu, X., Li, N., 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, Liangyan, Li, M., Dong, C., Ma, X., Wang, Lei, Song, D., Feng, J., Xu, P., Li, X., 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

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

Xu, J., Lin, S., Gao, X., Nie, Q., Luo, Q., Zhang, X., 2017. Mapping of Id locus for dermal shank

melanin in a Chinese indigenous chicken breed. J. Genet. 96, 977–983. https://doi.org/10.1007/S12041-017-0862-Z

Yan, L., Liqing, Z., Dexiang, Z., Xiquan, Z., Xuemei, L., 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., Wang, Y., Wei, X., Mao, L., Zeng, C., Wang, G., 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., Tang, M., Chen, J., 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., Zhao, D., Zhang, Y., Chen, W., 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

Zhu, Z., Ma, Y., Li, Y., Cheng, Z., Li, H., Zhang, L., Xu, D., Li, P., 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: A magnified view of a ~10.5 Mb (72000000-82529921) segment situated on the q-arm of Chicken chromosome Z: (A) Within this segment, two sets of Z amplicon repeat units (ZARU) which primarily harbor multiple copies of three genes: *C2orf3*, *ARHGAP33*, and *MRPL19*. In the initial part of the first Z amplicon repeat unit (ZARU), the *ADCY10* gene has

multiple copies. In contrast, towards the terminal end of chromosome Z, various copies of *ARHGAP33* are present, depicted in the initial panel. The second panel shows read coverage within a 10 Kb window of the same region and supports high coverage due to repeat regions. (B) Genes present in the R1 region have been shown. The direction of the arrows represents the direction of transcription.

Fig. 2: Phylogenetic tree and population structure analysis of the R1 region: (A) Phylogenetic tree based on SNPs (Single Nucleotide Polymorphisms) at the R1 region, showcasing 270 individuals (42 BBC and 228 non-BBC) and visualized using FigTree. **(B)** Principal Component Analysis (PCA) of the R1 region indicates that PC1 explains 19.8% of the genetic variance, while PC2 accounts for 9.01%. **(C)** An admixture analysis for individual ancestry reveals distinct separation of all BBC individuals at K6.

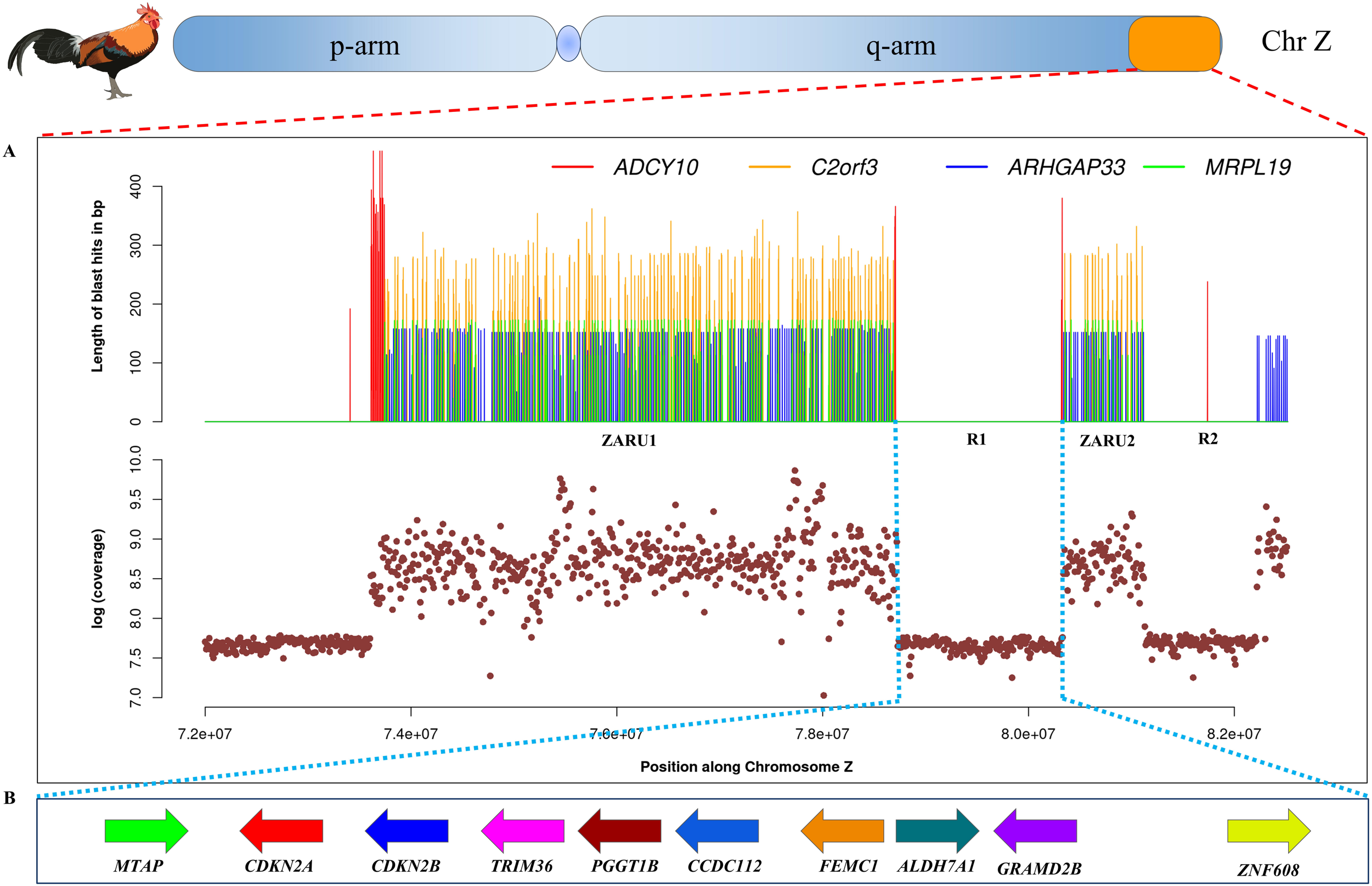
Fig. 3: Signature of selection in the last intron of the *MTAP* gene located in R1: (A) Comparative analysis of Weighted F_{ST} (KADK vs. WHLH + SILK vs. WHLH/KADK vs. SILK) between BBC (KADK and SILK) and non-BBC (WHLH). **(B)** F_{ST} comparison specifically between KADK and SILK. **(C)** F_{ST} comparison between KADK and WHLH. **(D)** FST comparison between SILK and WHLH. **(E)** Comparison of D_{xy} between KADK and SILK. **(F)** Comparison of D_{xy} between KADK and WHLH. **(G)** Further F_{ST} comparison between SILK and WHLH along chromosome Z, using 1Kb non-overlapping windows. A horizontal black dotted line represents the 90th percentile threshold for both F_{ST} and D_{xy} comparisons. **(H)** The phastCons conservation score of the GRCg6a genome assembly is displayed in dark blue, with a red dotted vertical line marking the selected region. **(I)** Identification of 6 SNPs located within a 5kb span in the last intron of the *MTAP* gene. The percentage of bases at each SNP has been documented for both the BBC (n=42) and non-BBC (n=228) breeds.

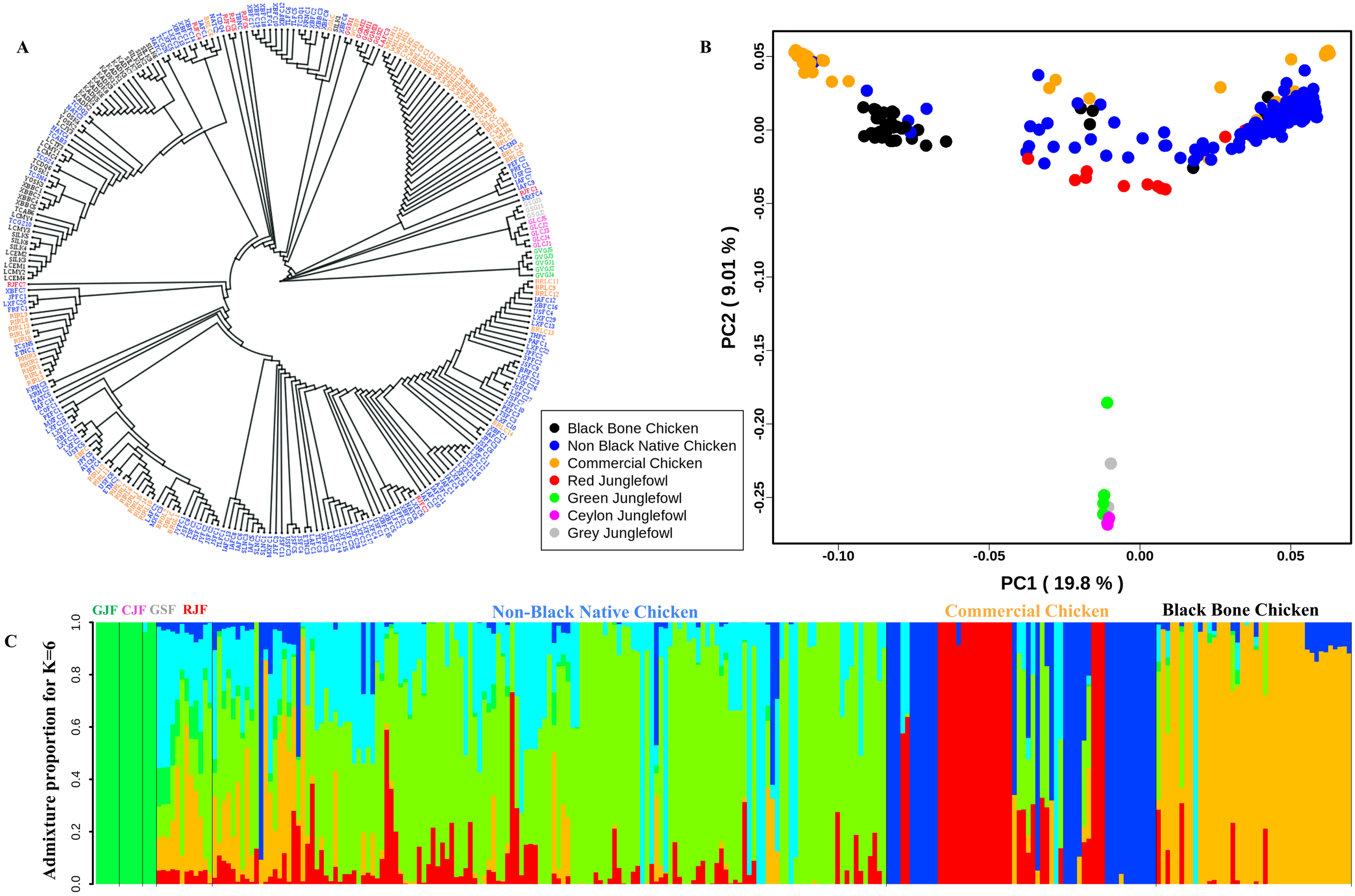
Fig. 4: Identification of three SNPs in the promoter region of the *CCDC112* gene: (A) Within the promoter region of the *CCDC112* gene, three distinctive SNPs, uniquely found in BBC, are illustrated by orange triangles. **(B)** This section presents the percentage of bases at these three SNPs for the BBC (n=42) and the non-BBC (n=228) breeds.

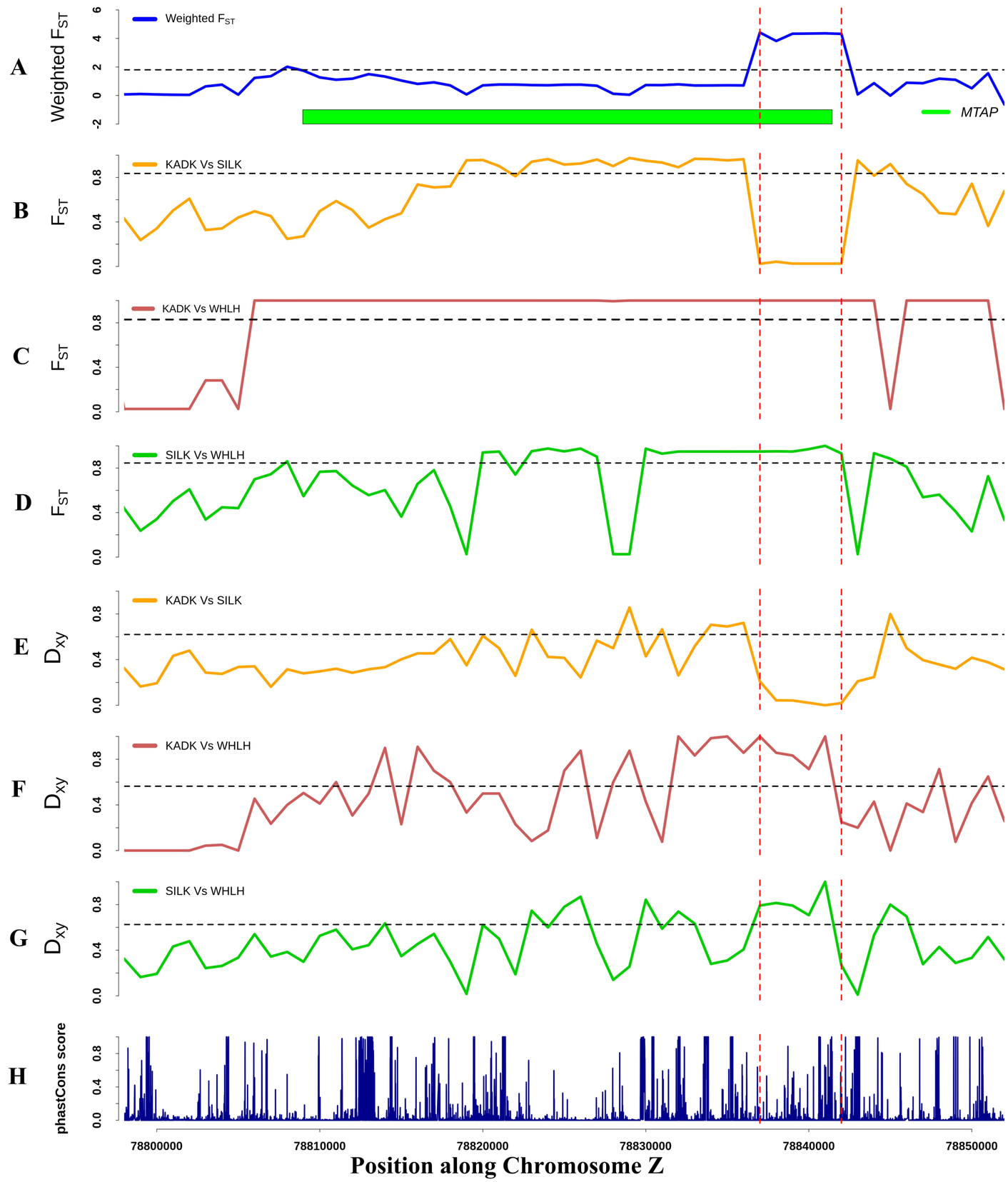
Fig. 5: Signature of selection in a 26 Kb long intergenic region located upstream to the ZARU: (A) Comparative analysis of Weighted F_{ST} (KADK vs. WHLH + SILK vs.

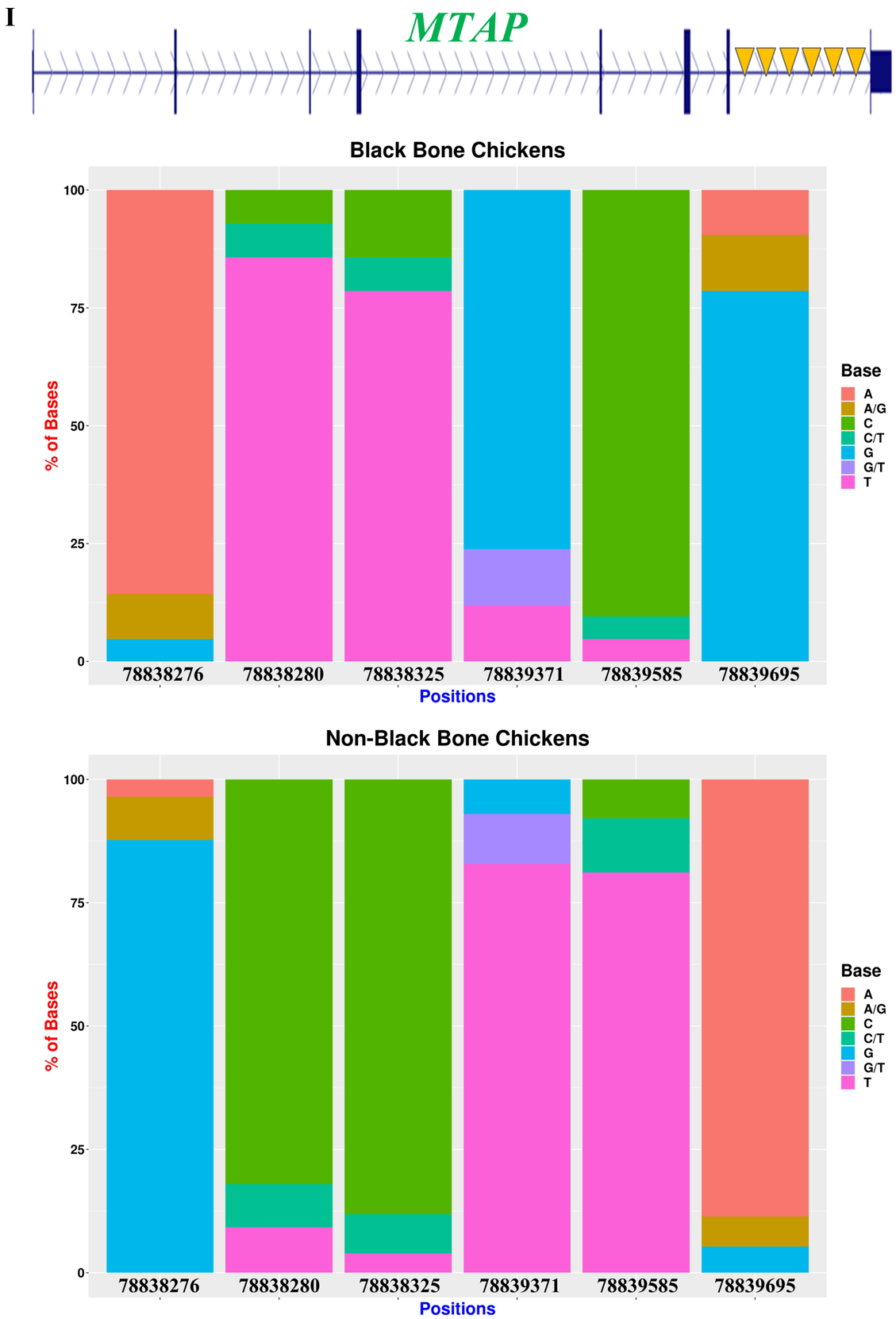
WHLH/KADK vs. SILK) between BBC (KADK and SILK) and non-BBC (WHLH). **(B)** F_{ST} comparison between KADK and SILK. **(C)** F_{ST} comparison between KADK and WHLH. **(D)** F_{ST} comparison between SILK and WHLH. **(E)** Comparison of D_{xy} between KADK and SILK. **(F)** Comparison of D_{xy} between KADK and WHLH. **(G)** Further F_{ST} comparison between SILK and WHLH, along chromosome Z, using 1Kb non-overlapping windows. A horizontal black dotted line represents the 90th percentile threshold for both F_{ST} and D_{xy} comparisons. **(H)** and **(I)** depict π and Watterson theta, respectively. Here, slate blue represents the KADK breed, firebrick represents SILK, and cyan represents WHLH. **(J)** The phastCons conservation score of the GRCg6a genome assembly is displayed in dark blue, with a red dotted vertical line marking the selected region. **(K)** and **(L)** presents the gene enrichment analysis for the target genes of gga-mir-1457 and gga-mir-6552.

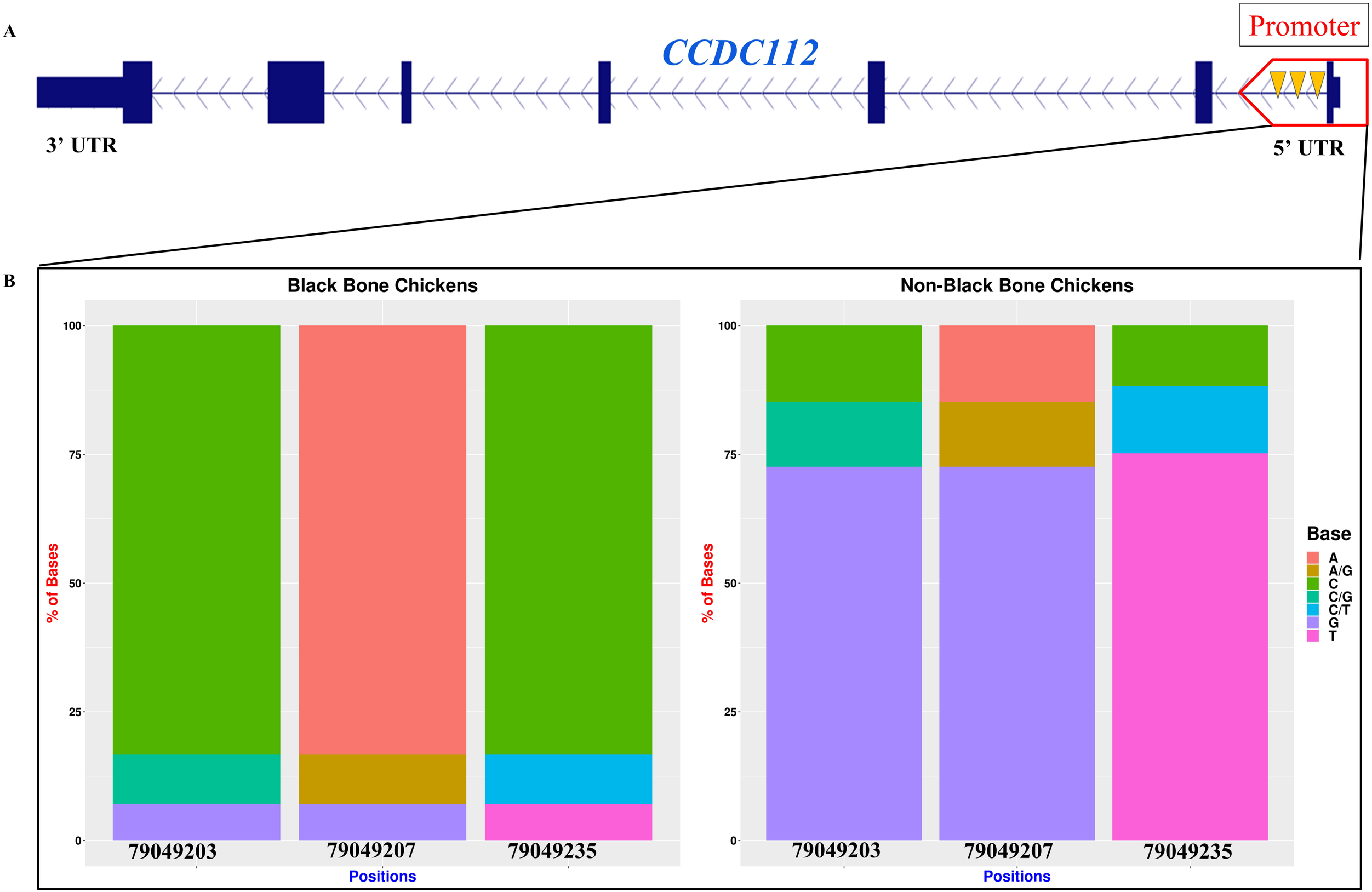
Fig. 6: Proposed novel genetic crossing design for identification of *Id* gene: A round-robin cross between a group of five breeds consisting of three BBC breeds (Kadaknath, Silkie, and Yeonsan Ogye) and two non-BBC breeds (White Leghorn and White Plymouth Rock).

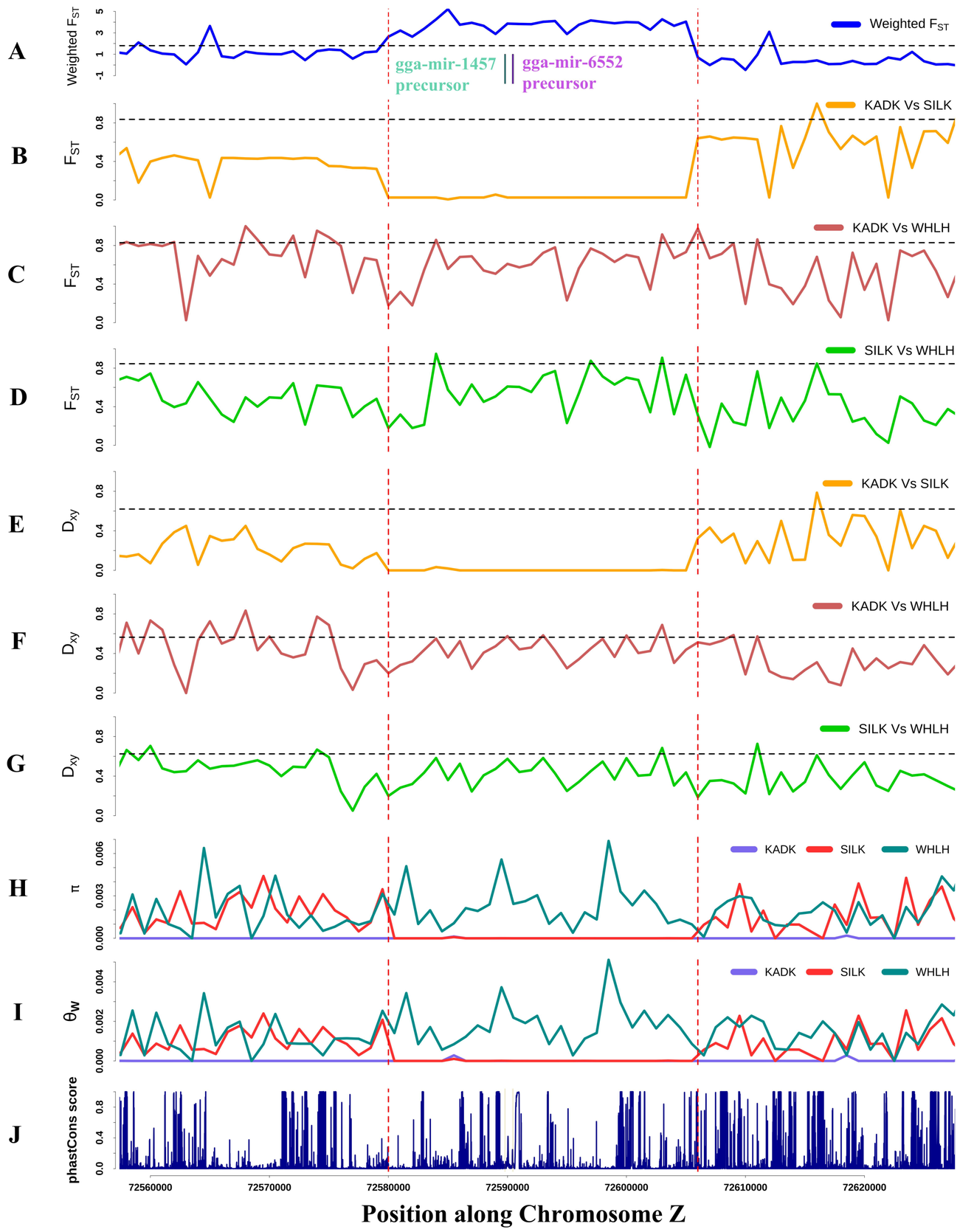

p-arm
q-arm
Chr Z
A
ADCY10
C2orf3
ARHGAP33
MRPL19
Length of blast hits in bp
0
100
200
300
400
ZARU1
R1
ZARU2
R2
log (coverage)
7.0
7.5
8.0
8.5
9.0
9.5
10.0
7.2e+07
7.4e+07
7.6e+07
7.8e+07
8.0e+07
8.2e+07
Position along Chromosome Z
B
MTAP
CDKN2A
CDKN2B
TRIM36
PGGT1B
CCDC112
FEMC1
ALDH7A1
GRAMD2B
ZNF608

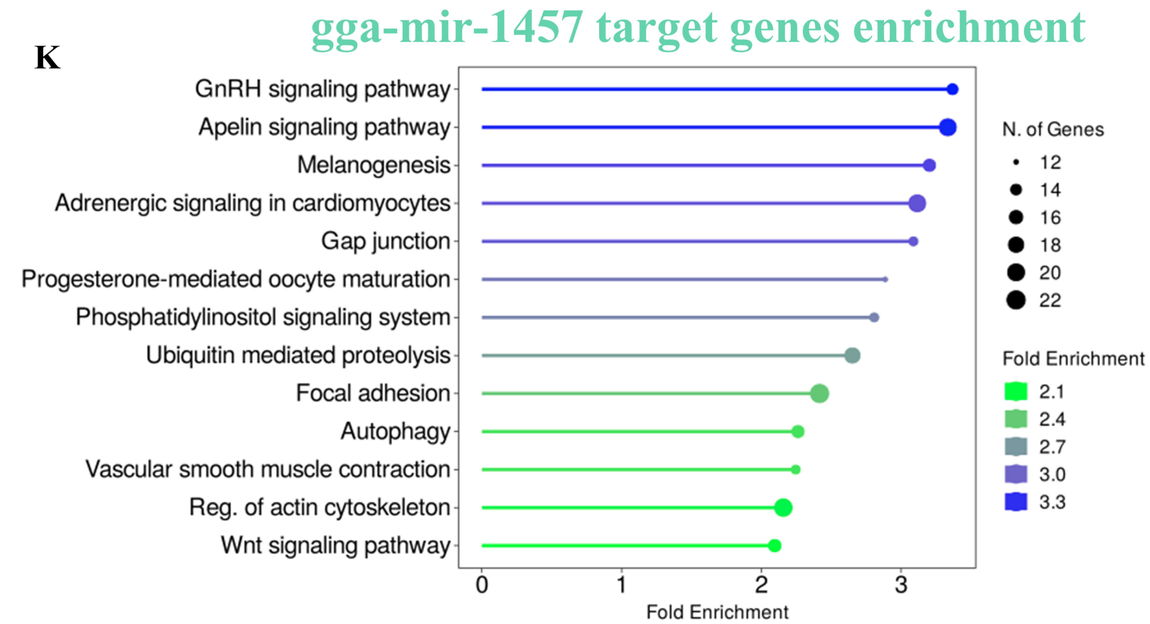

A
B
Black Bone Chicken
Non Black Native Chicken
Commercial Chicken
Red Junglefowl
Green Junglefowl
Ceylon Junglefowl
Grey Junglefowl
PC2 (9.01 %)
PC1 (19.8 %)
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-0.10
-0.05
0.00
0.05
C
GJF CJF GSF RJF
Non-Black Native Chicken
Commercial Chicken
Black Bone Chicken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

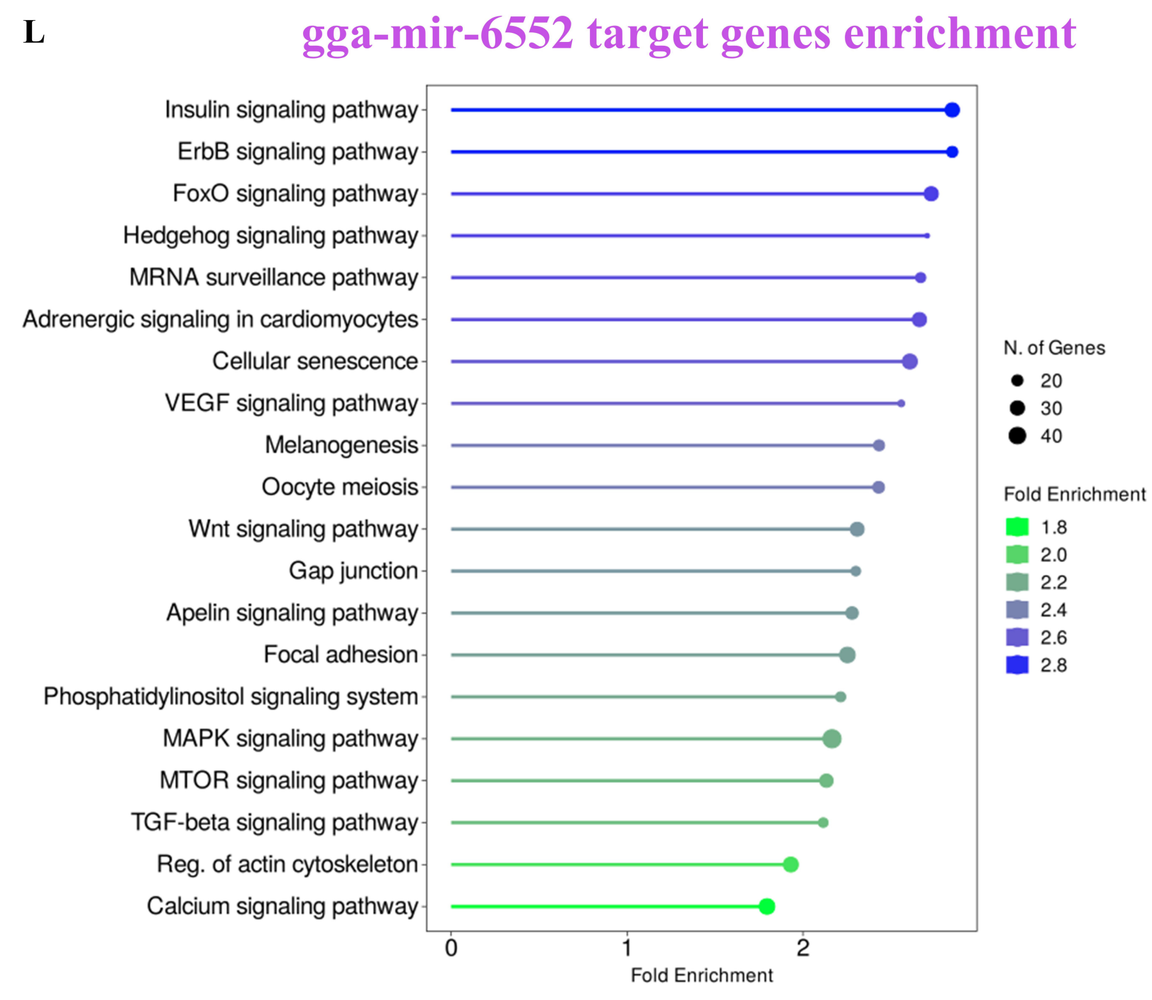

A
Weighted F_{ST}
Weighted F_{ST}
MTAP
B
F_{ST}
KADK Vs SILK
C
F_{ST}
KADK Vs WHLH
D
F_{ST}
SILK Vs WHLH
E
D_{xy}
KADK Vs SILK
F
D_{xy}
KADK Vs WHLH
G
D_{xy}
SILK Vs WHLH
H
phastCons score
78800000
78810000
78820000
78830000
78840000
78850000
Position along Chromosome Z

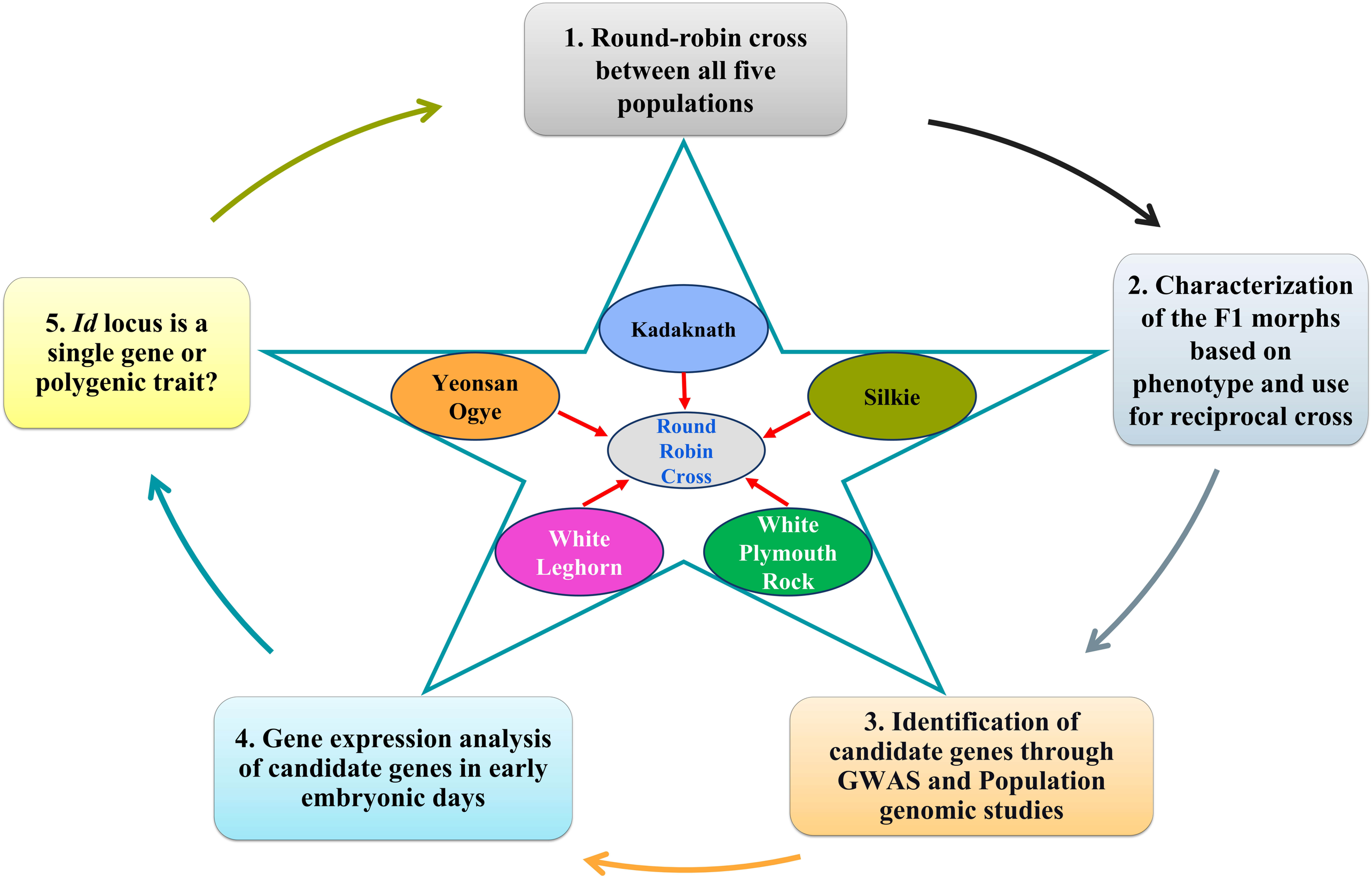

I
MTAP
Black Bone Chickens
% of Bases
Positions
78838276
78838280
78838325
78839371
78839585
78839695
Base
A
A/G
C
C/T
G
G/T
T
Non-Black Bone Chickens
% of Bases
Positions
78838276
78838280
78838325
78839371
78839585
78839695
Base
A
A/G
C
C/T
G
G/T
T


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
B
C
D
E
F
G
H
I
J
Weighted F_{ST}
gga-mir-1457 precursor
gga-mir-6552 precursor
KADK Vs SILK
KADK Vs WHLH
SILK Vs WHLH
F_{ST}
D_{xy}
KADK
SILK
WHLH
π
θ_W
phastCons score
72560000
72570000
72580000
72590000
72600000
72610000
72620000
Position along Chromosome Z


K
gga-mir-1457 target genes 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
12
14
16
18
20
22
Fold Enrichment
2.1
2.4
2.7
3.0
3.3


L
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
20
30
40
Fold Enrichment
1.8
2.0
2.2
2.4
2.6
2.8


1. Round-robin cross between all five populations
2. Characterization of the F1 morphs based on phenotype and use for reciprocal cross
3. Identification of candidate genes through GWAS and Population genomic studies
4. Gene expression analysis of candidate genes in early embryonic days
5. *Id* locus is a single gene or polygenic trait?
Kadaknath
Yeonsan Ogye
Silkie
Round Robin Cross
White Leghorn
White Plymouth Rock