Comparative population genomics reveal the genetic basis underlying feather color variation in domestic chickens

Ming-Yuan Zhu[1,#], Xing Guo [2,3,4,#], Shou-Qing Yan [1], Dong-dong Wu [2,3,4,*], Ming-Shan Wang[2,3,4*]

1. College of Animal Science, Jilin University, Changchun 130062, China
2. State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
3. Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China
4. Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture of China, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

Contribute equally to this work

* Corresponding author:

Dr. Dong-Dong Wu, email: wudongdong@mail.kiz.ac.cn

Dr. Ming-Shan Wang, email: wangmingshan@mail.kiz.ac.cn

Introduction

1-Domestication is an facilitating the transition of life style from hunt gather to settled. More than 40 species was successfully domesticated as having an essential rolling as food, biological models in medical research, phenotype evolution and developmental biology. A remarkable changing is this process is that domestic animals acquired enormous phenotypic diversity than their wild ancestors. One striking example is from chicken, after more than several thousand years breeding and domestication, chickens breeds were differs from each other in body size, feather type and color, crest type, etc., which is considered to have greatest phenotype diversity among birds.(1)

2-The feathers is the one of the most complex keratinized structure of the birds just as the hairs of mammals and has vital importance for concealment, sexual and species recognition at different stages of development.The complex organization of feather provides an excellent model for development and evolutionary biology as variation can occur at each step of differentiation and development. Feather display enormous diversity in domestic chicken, which differs in distribution (eg,whole naked, and necked, foot feathering),color(eg,brown,gray,black,white,green,pink,yellow or orange) and structure (eg,Frizzle, smooth-feather) among different breed. In recent years, researcher has made tremendous progress on research about feather,includeing that the Naked neck trait is caused by BMP12 gene which changes the distribution of feathers on the neck, the yellow pigment trait is caused by MuPKS gene which changes the accumulation of yellow pigments, and the Frizzle feather trait is caused by KRT75 gene which plays a significant role in characteristic curled feather rachis and barbs.

Domestic chicken has the huge variation in coloration compared to their wild ancestors through a long selective breeding history since the early days of domenstication ,which proven to be an excellent model organism for understanding the genetic of pigmentation considering plenty of pigmentation genes mapped in chicken.As for

carotenoid-based pigmentation,the yellow skin was cuased by the enzyme BCO2(beta-carotene oxygenase 2) which allow the depositions of colorful carotenoids carotenoids. Research in the field of melanin-based pigmentation biology is extensive,and the pathway associated with the production of pheomelanin pigment involoved some genes like ASIP,MC1R,Tyr,Tyrp1,Tyrp2 that affect my work.

Melanins, porphyrins, sprains, carotenoid and polyenes have all been observed in feather that showed a very striking and intricate color pattern, but the specific causative genes haven't been characterized.(2) The pioneer studies were largely with commercial or homozygosity chickens containing little of variation and their results could be better convinced if they consider that inheritance of chicken feather color is deceived by a handful of genetic factors ,which cause plenty of troubles for both the biologist and breeding researcher.

3- However, the emergence of NGS has revolutionize the research in evolution and genetics. Complex trait, like seasoning reproduction, highland adaptation, body size variation, and vision, have been elicited, which greatly facilitate our understanding of genetic changes in their evolution. In this research, taking the advantages of NGS, we try to using population genomic to discern potential genes determine the color variation in domestic chickens, potentially in for the genetic basis underlying feather color variation.Yuanbao chicken, an ornamental famous chicken variety which is famous for its gold ingot appearance, has a long breeding history from the Tang Dynasty for hundreds of years. Yuanbao chicken is excellent biological models for its specific phenotype like small body size and solid-colored feather thus been widely used in development and evolution research. To promote a better understanding of the variants affecting feather color changes in the domestic animals, we use the comparative population genomics based on the NGS to study the genome of Yuanbao chickens . We identified four novel loci that potentially control the variation in color pattern of domestic chickens,which will provide new insight about basis pigmentation biology.

Results

Genetic diversity

Summary of mapping, SNPs calling, PI of each chromosome and density

In the present study, 24 genomes were obtained representing 7 Yuanbao white chickens, 7 Yuanbao black chickens, and 7other domestic chickens. Within the genic regions, the proportion of SNPs occurring in intergenic,intronic,up down stream,ncRNA,exonic and UTR regions was 9112159, 7584082, 552045, 488986, 287323 and 52759, respectively.(Figure 1A) Functional annotation of the SNPs assigned to protein-coding regions identified 85706 SNPs that produce non-synonymous amino acid substitutions and 200747 SNPs that were synonymous, with 870 genes having SNPs that cause gain or loss of a stop codon.

To evaluate heterozygosity at the population level, nucleotide diversity (π) was calculated as the number of heterozygous individuals relative to homozygous individuals at all SNP sites in sliding windows of the 24 genomes.

Distribution of nucleotide diversity, measured as Tajima's D, across sliding windows by using a 50-kb sliding window with 25-kb stepwise were also calculated throughout the genome. (Figure 1B), the data are summarized for each chromosome across all windows.Then the density of SNP per 100 kb of genic on each chromosome were calculated. Overall and for comparison, we observed the highest number of SNPs on chromosome 23 and the lowest number of SNPs on chromosome W, and the number of SNPs was correlated with the length of the chromosomes. But the boxplot show that W chromosome diversity is lower than expected even when differences in effective population size are taken into account. The observations of lower-than-expected levels of genetic variability in the Y chromosome of organisms with male heterogamety and in the W chromosome of the domestic chicken with female heterogamety indicate that sex-limited chromosomes generally possess limited genetic diversity. Nonrecombining chromosomes are susceptible to the influence of selective forces acting on any sequence on the chromosome, and given that selection is the only factor broadly applicable to explain reduced diversity in the sex-limited chromosome, irrespective of mode of reproduction, or whether there is male or female heterogamety, we

argue that selection may be a common denominator of reduced variability in Y and W sex chromosomes.~~**Genetic diversity**~~

~~Xxx howmany SNP；~~

~~How is the distribution~~

~~Pi for each chromosmo~~

An approach of comparative population genomics to ascertain loci controlling chicken feather color

Under domestication, domestic chicken has swiftly evolved genetic adaption to the environment underlying phenotypic changes. Based on the genomic variation data acquired before, we characterized candidate genes that potentially contribute to the different color phenotypes between yuanbao black and white chickens.

Based on the processed data, We employed a comparative analysis of population variants to ascertain candidate selected genes. This powerful tools is considered to be an effect of a method into genetic which are underlying complex traits. (Evolution of Darwins clustered in genomic loci and biological pathway affect human height) Generally by going through selection, loci or genomic regions will evolve swiftly which would display specific signatures of variation, including high population differentiation, significantly reduced nucleotide diversity levels and long-range haplotype homozygosis. Based on these principles, we examined three different parameters to identify footprints of artificial selection associated with the evolution and domestication of Yuanbao chicken from other chicken: Fst, nucleotide diversity (Pi). LSBL. The first percentile rank was used as a threshold to identify candidates throughout the analysis. First ,a sliding window analysis was performed ,with 50-kb window size and 25-kb step size ,identifying genes with significant higher FST (403 genes),LSBL(b-w-o 369genes,w-b-o 368 genes) and a lower value for nucleotide diversity (Pi, 403 genes),respectively ,as candidates based on the outline approach (99th percentile cutoff).Functional enrichment analysis of these candidate genes did not reveal any pathway specifically associated with the development of body size(Supplementary Table).By combining the signals of these approach, we

identified three regions of the genome (chr1: 187.30Mb-188.50Mb, ch1:191.45Mb-191.62Mb, ch2:69.1Mb-72.9Mb) that exhibited extreme population differentiation,likely as the result of artificial selection.

Analysis of chr1:187.30Mb-188.50Mb shows that Tyr and Rab38 potentially controlling the feather color in black and white Yuanbao chicken

The genomic region ch1:188.25Mb-188.50Mb stands out as the most extremely candidate selective sweep with the highest level of population differentiation.There are five high score genes: CTSC,TYR, RAB38,gga-mir-1657, and mGluR5a located in this region.

CTSC associated with molecular function like cysteine-type endopeptidase activity and serine-type endopeptidase activity indicating that CTSC may have no dependency with the pigment. TYR gene, codes the tyrosinase infecting the biosynthesis of melanin which is the key enzyme in melanin biogenesis in pigment cells. The melanin synthesis pathway will be blocked if the enzymatic function of tyrosinase is abnormal, resulting in an albino phenotype. The chicken tyrosinase (TYR) gene has everal mutations at the C locus (Smyth J. R. Ring N. M. Brumbaugh J. A. 1986. A fourth allele at the C locus of the chicken. Poult. Sci. 65(Suppl. 1):129. (Abstr.)) include the recessive white mutation that has the insertion of a complete avian retroviral sequence in intron 4 of the TYR gene, which becomes the diagnostic characteristic of the recessive white mutation.

RAB38, a member of the Rab family proteins, showed consistently higher values in both b_w_o LSBL and w_b_o LSBL analyses.Until now, a role of RAB38 in feather color has not been reported in chickens. gga-mir-1657, a non-coding RNAs (ncRNAs) predicted using sequences from RFAM and miRBase, resides within host gene RAB38. Specific Rab proteins such as Rab38, have been shown to play a key role in in melanosome biogenesis. Further investigation showed that RAB38 had a high expression in the skin of mice, but there is no experiment of it in the chicken feather. An analysis of coat color mutant “chocolate” identified the Rab38 gene involved in the regulation of pigmentation and played a role in the sorting of TYRP1 which participated in melanin synthesis(15).This suggests that Rab38 regulate a

critical step in the trafficking of melanogenic enzymes, in particular, tyrosinase, from the TGN to melanosomes.RAB38 participates in endosome to melanosome transport and melanosome Assembly,which confirm our previous view in some part.

RAB38 and Tyr gene both involved in the accumulation of pigment in an organism, tissue or cell, either by increased deposition or by increased number of cells （GO:0043473）, constitute a small, subcellular membrane-bounded vesicle containing pigment and/or pigment precursor molecules（GO:0048770）as well as participated in the construction of melanosome（GO:0042470）.mGluR5a (glutamate metabotropic receptor 5), plays vital roles in neuronal development glutamate receptor activity and G-protein coupled receptor activity involved in regulation of post synaptic membrane potential.It is likewise acting as a glial sensor of the extracellular glutamate concentration in order to acutely regulate the excitatory transmission.(Acute Up-regulation of glutamate uptake mediated by mGluR5a in reactive astrocytes.)

Genes at ch1:191.45-191.62, ch2:69.1-72.9 Mb are potentially involoved in the differences of feather color

The observation that ch1:191.45-191.62 and ch2 69.1-72.9 Mb regions have the high score in both b_w_o and w_b_o LSBL raised the possibility that this region may affect selection. Ch1:191.40Mb-191.52Mb region contains 215 protein-coding genes. There is no QTL associated with feather color or skin color in this region. These 215 genes are involved in various biological procession. For example,UVRAG is associated with retrograde vesicle-mediated transport, while PAAF1 are involved in proteasome complex.

Third mapped region ch2:69.1Mb-72.9Mb, harbouring a cluster of selective sweep SNPs,demonstrates significantly higher levels of population differentiation as revealed by W_B_O LSBL and B_W_O LSBL. The gene MC4R(melanocortin 4 receptor), which participates in melanocortin receptor and melanocyte-stimulating hormone receptor activity. The melanocortin pathway is involved in the regulation of several physiological functions including skin pigmentation, steroidogenesis,

obesity, energy homeostasis, and exocrine gland function. MC4R located in the brain are implicated as participating in the metabolic and food intake aspects of energy homeostasis(18) In addition ,no QTL associated with feather in chickens was mapped in this gnomic region.

Numbers of genes identified by the three method listed in each of the Venn diagram components

To determine whether the three approaches have consistency of conclusion, we performed Venn Diagram. As showed in figure,only 27were identified by all three methods (Figure 1C,). It is not surprising that the overlap among the positively selected genes (PSGs) detected by the different statistical approaches was underwhelming. ~~First, distinct statistics applied to scan the genome for positive selection are based on different signatures of population variation. Another reason might be attributable to a pitfall in the outlier approach, whereby an outlier value in one analysis, falling in the 99th percentile of the empirical distribution, may not be classified as an outlier in another study where it may fall within the 98th percentile of the study's empirical distribution. This pattern is quite common in studies on positive natural selection in humans. For example, only 14.1% loci were identified in two or more studies with large-scale genome-wide scans of PSGs. In the current study, possibly due to the above stated factors, no significant enrichment for any functional category associated with vision was found among the candidates identified by the FST test~~. A number of protein-coding genes with significantly higher score in the intersection have been identified.

For example,except for the genes included in the regions that we discussed before,we identified two genes, PARN and F2RL1, both of which were previously reported participating in pigments disease in human. PARN, at Chromosome 14: 0.793Mb-0.828Mb associated with dyskeratosis congenita (DC), a telomere biology disorder, is characterized by a classic triad of dysplastic nails, lacy reticular pigmentation of the upper chest and/or neck, and oral leukoplakia.（Dyskeratosis Congenita）PARN related to Dyskeratosis congenita (DC), a telomere biology

disorder, is characterized by a classic triad of dysplastic nails, lacy reticular pigmentation of the upper chest and/or neck, and oral leukoplakia.（Dyskeratosis Congenita）Similar ,the gene F2RL1,which located at Chromosome Z: 23,413,421-23,422,623,（17）indicated in the experiment that F2RL1 variants are important for pigmentation as well as for melanoma risk and G-protein coupled receptor activity ,also had the highest identified score in both three approach by the sliding window analysis.F2RL1 have also been reported previously（17）to be of great importantance for pigmentation as well as for melanoma risk.

Discussion

Chickens are of great importance in biological research and entertainment, and has been used for quantitative genetic studies for decades. For the source of various products including pillow stuffing, insulation, upholstery padding and bio-diesel as well as other advantages like ornamental value and convince for processing of carcass, breeders have made great effort to select the pure color chicken which is homozygous dominant. Under domestication, domestic chicken has swiftly evolved genetic adaption to the environment underlying phenotypic changes. Based on the genomic variation data acquired before, our study reports a serious of candidate genes that potentially contributes to the different color phenotypes between Yuanbao black and white chickens and candidate genetic marks for chicken breeding. Our study also provides a strategy with comparative population genomics to open a door for identifying candidate genes/variants accounting for the variation in feather color of chickens. This strategy is much more cost-effective and timesaving than previous methods such as QTL mapping and GWAS analysis.

A great variation in feather color has been observed in several domesticated animals, including parrots, guinea pig and chicken. Feather color is not only an important trait for commercial production,but also a key topic for evolutionary and developmental biology studies.The investigation in genes/genetic variants controlling variation in feather color arose a fierce focus from animals breeders,evolutionary biologists,and even medical scientists.

Feather color, typical of many complex traits, is commonly believed to be influenced by many genes involved in similar function pathways. TYR, TYRP1 and SLC24A5 have been previously shown to influence melanin-based color patterns in vertebrates.In assessing the genetic underlying the feather color of Yuanbao chicken, we identified several genomic regions associated with melanin genes.In a selective sweep region in chromosome 1 (ch1:187.30-188.50Mb),there were 6 protein-coding genes with diverse biological functions. Tyrosinase (TYR) is the key enzyme is the rate-limiting enzyme in the production of melanin pigment, causing different plumage color patterns in birds: gain-of-function mutations lead to the synthesis of eumelanin, whereas loss-of-function mutations help to generate pheomelanin synthesis. Tyr gene has also been referred to as a textbook example of researching white and black plumage formation to explain melanic variation among both chicken and other vertebrate. RAB38 results in the feather color variation remains to be determined, but has been identified related to melanin biogenesis and involved in the regulation of pigmentation. Since recent data support that Rab38 have an important role for sorting the TYRP1 which related to melanin production and participate in endosome in the melanin in some transport.
gga-mir-1657, which host gene is Rab38, is a non-coding RNAs (ncRNAs) in this region, and it's position suggest that it may be important for the modulation of Rab38 gene expression. Our finding that the gene of this region involved in the accumulation of pigment , that participated in the construction of melanosome and other biological process of melanin and that has been strongly selected by domestication ,collectively suggest that these genes may play a role in creation of different feather color.

Accordingly, we consider the possibility that the Rab38 ,Tyr and other genes in this region cause the main influence of the variants between Yuanbao black and white chickens.In this work, we have presented evidence implicating that the genes in this region are in a novel signaling pathway that regulate differentiation and melanin production in Yuanbao chicken.

We also found other important genes such as UVRAG, PAAF1 at ch1:191.40-191.52Mb,MC4R at ch2 :69.1-72.9Mb PARN at Chromosome 14: 793,82Mb and F2RL1 at Chromosome Z: 23,41Mb, that associated with various of biological process. MC4R(melanocortin 4 receptor) is the melanocortin receptor and melanocyte-stimulating hormone receptor activity.The melanocortin pathway is involved in the regulation of skin pigmentation.UVRAG is associated with retrograde vesicle-mediated transport,and PAAF1 are involved in proteasome complex.F2RL1 and PARN are of great importance for pigmentation and other skin disease like Dyskeratosis Congenita.Nowadays,numerous studies suggest that it has a preventive role in pigmentation and transshipment of melanin,although the mechanism of action still remains unclear and mounting evidence may suggest that they are participated in all kinds of biological process related to melanin.Surprisingly,this results established a link between feather color and chromatosis of vertebrate represented by human. The findings offer a new insight into the mechanism by which.

The above findings raised the intriguing possibility that some genes undergoing a selective sweep as well as present a high score in LSBL but further investigations shows they have a little dependence with feather color maybe for the genetic hitchhiking affection. We found that coding sequence mutation were slightly enriched in the putative selective sweeps. This observation suggests that physiological traits are likely to be artificially selected science adaptive mutations affecting physiology are more likely to appear in the protein-coding regions.

Materials and methods

Assembling genomic analysis data

All 24 chickens genomes were culled from our preceding study consisting of 8 black Yuanbao chickens, 8 white Yuanbao chickens, as well as 8 other domestic chickens(9).

Genomic sequence alignment and genotyping

We cleaned raw high-throughput sequencing reads by removing low-quality bases ,adaptors, primers, and other types of useless sequence by using Btrim and cutadaptor software (10). By using BWA-MEM with initial settings, filtrated pair-end reads were mapped against chicken reference genome (Galgal4) (aligning sequence reads, clone sequences and assembly contigs with BWA-MEM). A battery of post-processes were then utilized to process the alignment BAM format file, involving sorting, duplicates marking, local realignment, and base quality recalibration, which were carried out using the SortSam and MarkDuplicates functions in the Picards (picard-tools-1.56, http://broadinstitute.github.io/picard/) package,and RealignerTargetCreator, IndelRealigner, and BaseRecalibrator tools in the Genome Analysis Toolkit (GATK) (11). SNPs and indels were called and filtered using UnifiedGenotyper and VariantFiltration command in GATK.Loci with the more than 2 alleles were removed. All SNPs were assigned to specific genomic regions and genes utilizing ANNOVAR based on the ENSEMBL chicken annotations (12). Population variation and population genetic analyses Genome-wide genetic diversity (π) was calculated for Yuanbao chicken and other chicken groups using VCFtools (13) using a 50-kb sliding window with 25-kb stepwise increments.

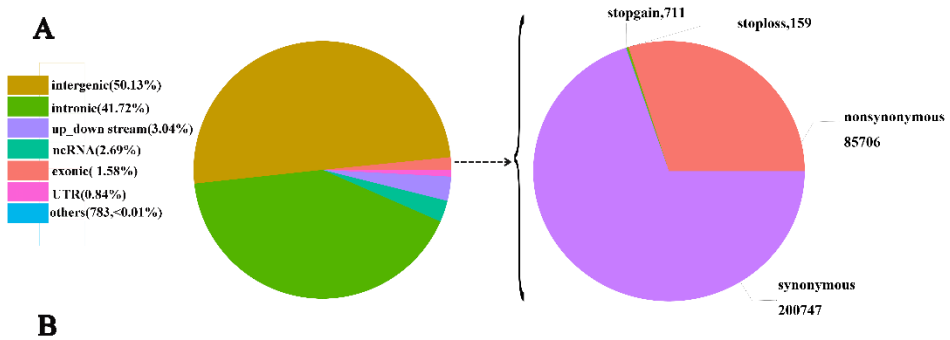
Genome-wide Scan for Selective Sweeps

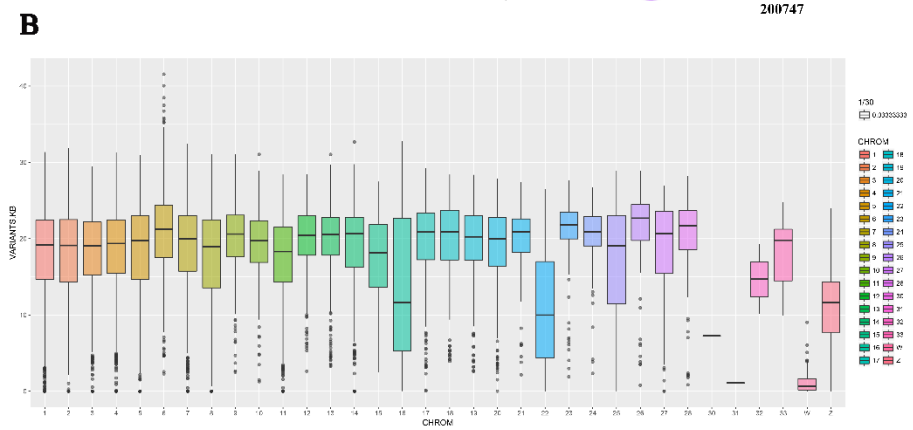
To determine genomic regions harboring footprints of positive selection in Yuanbao chicken, we applied multiplex tests to investigate selection in populations. Here we used the FST, LSBL, and Pi statistical methods. SNP FST values were calculated between Yuanbao black and white chickens as previously mentioned (14). LSBL statistics were computed for each SNP relied on the FST values between the three groups (15). So we defined Yuanbao black chicken as Group A, Yuanbao white

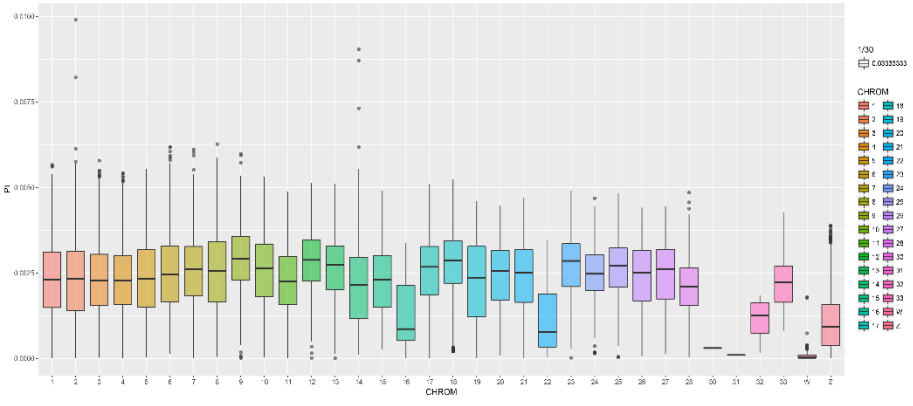
chicken as Group B., other domestic chicken lines were assigned group C(b_w_o LSBL). LSBL statistics for each variant were calculated using the formula: LSBL = (FST(AB) + FST(AC) - FST(BC))/2. Then we also use the Yuanbao white chicken as Group A, Yuanbao black chicken as Group B, other domestic chicken lines were assigned to group C to repeat the steps above(w_b_o LSBL).Sliding window analysis was performed for LSBL in each 50-kb window with 25-kb stepwise increments. Nucleotide diversities (ΔπorΔPi) = πRJF − πVC were calculated by using a sliding window analysis with a window size of 50 kb and a step size of 25 kb.

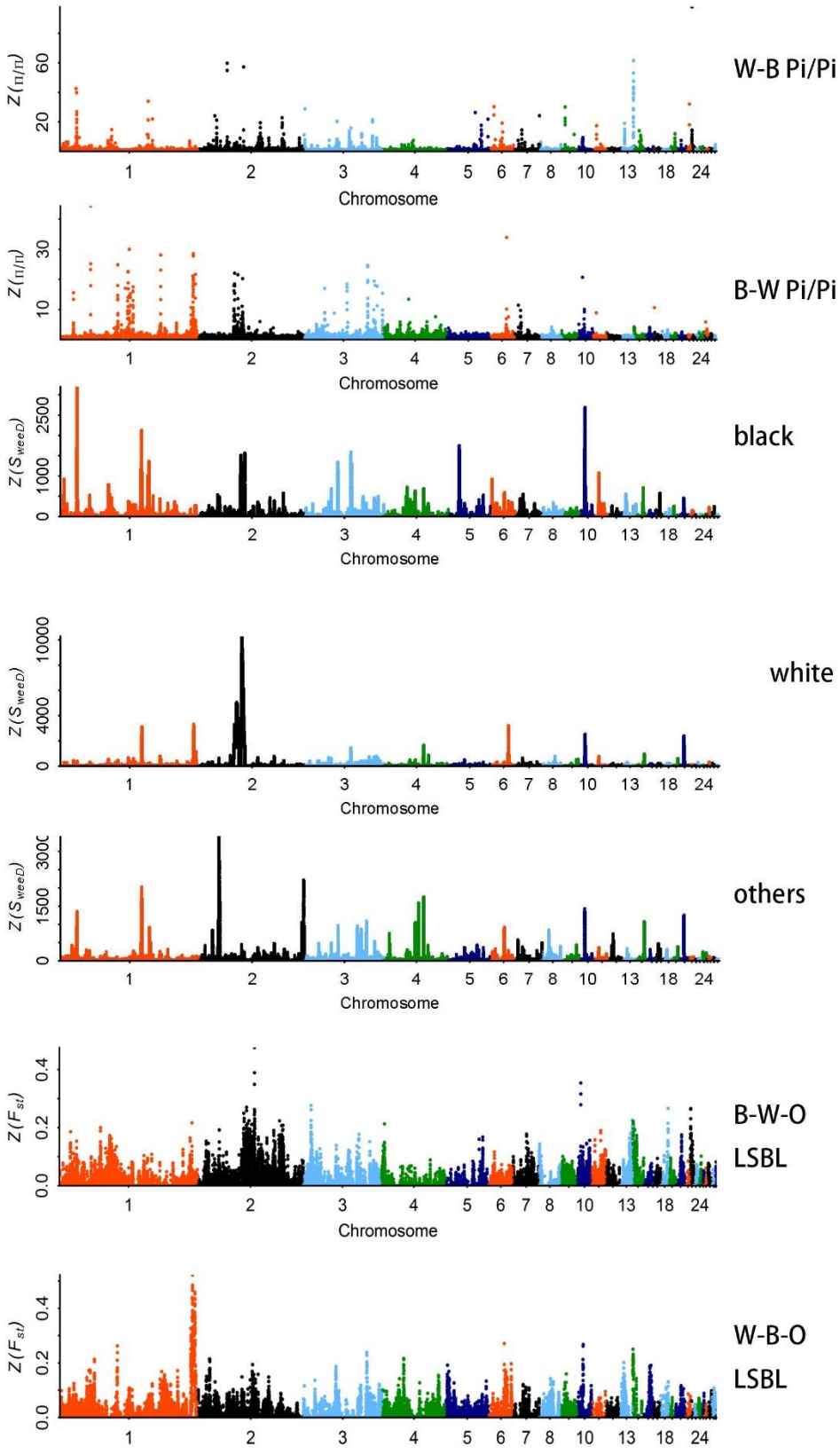
Selective sweep region annotation and gene functional enrichment analysis

Candidates' selective sweeps detected by the above-mentioned methods were annotated using the Variant Effect Predictor available at http://asia.ensembl.org/info/docs/tools/index.html. Functional enrichment of the protein-coding genes including GO categories, KEGG pathway and HPO were analyzed using g: Profiler(http://biit.cs.ut.ee/gprofiler/index.cgi) as well as DAVID(DAVID Bioinformatics Resources 6.8, at https://david.ncifcrf.gov/).


A
intergenic(50.13%)
intronic(41.72%)
up_down stream(3.04%)
ncRNA(2.69%)
exonic(1.58%)
UTR(0.84%)
others(783,<0.01%)
stopgain,711
stoploss,159
nonsynonymous
85706
synonymous
200747


B
VARIANTS/KB
CHROM
1/30
CHROM


C
PI
CHROM
1/30
CHROM


Z(π/π)
20
60
1 2 3 4 5 6 7 8 10 13 18 24
Chromosome
W-B Pi/Pi
Z(π/π)
10
30
1 2 3 4 5 6 7 8 10 13 18 24
Chromosome
B-W Pi/Pi
Z(SweeD)
0
1000
2500
1 2 3 4 5 6 7 8 10 13 18 24
Chromosome
black
Z(SweeD)
0
4000
10000
1 2 3 4 5 6 7 8 10 13 18 24
Chromosome
white
Z(SweeD)
0
1500
3000
1 2 3 4 5 6 7 8 10 13 18 24
Chromosome
others
Z(Fst)
0.0
0.2
0.4
1 2 3 4 5 6 7 8 10 13 18 24
Chromosome
B-W-O
LSBL
Z(Fst)
0.0
0.2
0.4
1 2 3 4 5 6 7 8 10 13 18 24
W-B-O
LSBL