



Patterns of Geographical and Potential Adaptive Divergence in the Genome of the Common Carp (*Cyprinus carpio*)

Jian Xu^{1†*}, Yanliang Jiang^{1†}, Zixia Zhao¹, Hanyuan Zhang¹, Wenzhu Peng², Jianxin Feng³, Chuanju Dong⁴, Baohua Chen², Ruyu Tai¹ and Peng Xu^{2,5*}

¹ Key Laboratory of Aquatic Genomics, Ministry of Agriculture, CAFS Key Laboratory of Aquatic Genomics and Beijing Key Laboratory of Fishery Biotechnology, Chinese Academy of Fishery Sciences, Beijing, China, ² State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen, China, ³ Henan Academy of Fishery Science, Zhengzhou, China, ⁴ College of Fishery, Henan Normal University, Xinxiang, China, ⁵ Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine Science and Technology, Qingdao, China

OPEN ACCESS

Edited by:

Paulino Martínez,
University of Santiago de
Compostela, Spain

Reviewed by:

Shaojun Liu,
Hunan Normal University,
China
Fabyano Fonseca Silva,
Universidade Federal de Viçosa,
Brazil

*Correspondence:

Jian Xu
xuj@cafs.ac.cn
Peng Xu
xupeng77@xmu.edu.cn

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 13 December 2018

Accepted: 24 June 2019

Published: 12 July 2019

Citation:

Xu J, Jiang Y, Zhao Z, Zhang H,
Peng W, Feng J, Dong C, Chen B,
Tai R and Xu P (2019) Patterns of
Geographical and Potential Adaptive
Divergence in the Genome of the
Common Carp (*Cyprinus carpio*).
Front. Genet. 10:660.
doi: 10.3389/fgene.2019.00660

The common carp, *Cyprinus carpio*, is a cyprinid fish species cultured in Europe and Asia. It accounts for >70% of freshwater aquaculture production worldwide. We conducted a population genomics analysis on *C. carpio* using high-throughput SNP genotyping of 2,198 individuals from 14 populations worldwide to determine the genetic architecture of common carp populations and the genetic bases for environmental adaptation. Structure analyses including phylogeny and principal component analysis were also conducted, showing distinct geographical patterns in European and Asian populations. The linkage disequilibrium block average lengths of the 14 populations ranged from 3.94 kb to 36.67 kb. Genes within selective sweep regions were identified by genome scanning among the different populations, including *gdf6a*, *bmpr1b*, and *opsin5*. Gene Ontology and KEGG enrichment analyses revealed potential trait-related loci and genes associated with body shape, scaling patterns, and skin color. This population genomics analysis may provide valuable clues for future genome-assisted breeding of *C. carpio*.

Keywords: common carp, population genomics, linkage disequilibrium, haplotype, selective sweep

INTRODUCTION

The common carp, *Cyprinus carpio*, is one of the most important cyprinid species due to its food value and complex paleotetraploidized genome. It is cultured in >100 countries, and the annual global production of *C. carpio* is > 4.56 million metric tons. This is approximately 10% of the global freshwater aquaculture production (FAO Fisheries and Aquaculture Department; Bostock et al., 2010). Common carp provide high value protein as food and some strains, such as koi, are popular ornamental fish. Common carp have been cultured for several thousand years. Domesticated common carp differ from their wild ancestor in morphological, behavioral, physiological, and reproductive traits (Balon, 1995). For example, wild carp has an elongated body with full scale cover, while domesticated carp usually have a much deeper body with four scale patterns: 1) leather, with no scales; 2) line, with large scales along the lateral line; 3) mirror, with a small number of large scattered scales; and 4) fully scaled. Domesticated carp are generally more capable of coping with extreme environments than their wild ancestor (Balon, 1995). Genetic evidence indicates

that all common carp populations originate from two ancestral forms of wild carp, the European subspecies (*C. c. carpio*) and the East Asian subspecies (*C. c. haematopterus*) (Chistiakov and Voronova, 2009). The validity of a third subspecies, *C. c. rubrofasciatus*, is questionable; it may have diverged from *C. c. haematopterus* (Zhou et al., 2004; Wang et al., 2010; Kohlmann and Kersten, 2013).

During its domestication, common carp has been introduced into many areas. Common carp ancestors have been subjected to genetic interventions, and natural and artificial selection. These factors combined with accumulation of mutations and long-term geographical isolation have produced many varieties of common carp with distinct skin color, body shape, scale pattern, body size, and stress tolerance. Human transport of carp to different geographical locations has generated high levels of gene flow (Wang et al., 2010). Hybrid breeding of carp in China has been common over the last 50 years and has resulted in many varieties or strains, such as Jian carp (JIAN). Multiple rounds of hybridization and genetic introgression were employed during hybrid breeding (Dong et al., 2015). Consequently, the genetic backgrounds of most common carp populations are unknown, especially when breeding history is inadequately recorded or missing. Various genetic tools and molecular markers have been developed and used for studying the phylogenetic relationship among populations and the genetic architecture of populations. These include random amplified polymorphic DNA, amplified fragment length polymorphism, restriction fragment length polymorphism, mitochondrial DNA, and microsatellites (Bartfai et al., 2003; Zhou et al., 2003; Mabuchi et al., 2008; Cheng et al., 2010). However, due to the limited resolution of these genetic markers, many phylogenetic relationships remain uncertain. For example, the origin and relationships among Hebao carp (HB), Xingguo Red carp (XG), Songpu carp (SP), Oujiang color carp (OUJ), and Koi carp (KOI) are controversial (Balon, 1995; Froufe et al., 2002; Wang and Li, 2004a; Wang and Li, 2004b). The previous studies indicated different phylogeny patterns using both nuclear genome (Xu et al., 2014) and mitochondrial genome sequence (Dong et al., 2015; Liu et al., 2019); however, due to limited number of samples, these results were not so solid for validation. Also, the genomic basis of local adaptation shaped by natural selection is still largely unknown. Selective sweep analysis is an effective approach to identify trait-related genes under natural selections or domestications. Xu et al. have reported the selective sweeps in the HB population compared with the SP population and identified *fgfr1a1* in the selective regions (Xu et al., 2014). Another research on Amur ide alkaline adaptation was also conducted using selective sweep method, and dozens of ion transportation-related genes were revealed involving in osmoregulation and pH regulation (Xu et al., 2017). Larger samples would be more persuasive in identification of genes in selective sweep regions due to the high genetic diversity of the populations.

With the fast growth of sequencing technologies, high-throughput genetic markers, such as single-nucleotide polymorphism (SNP), have been used in population genetics. Many studies have demonstrated that SNP arrays can improve the resolution of the differentiation of genetic stocks

(Perez-Enriquez et al., 2018; Torati et al., 2019). SNP assays are a useful tool for studying population structure and the effects of natural and artificial selection at the genome scale. For example, the Atlantic salmon SNP array that contained 6,176 informative SNPs was used to genotype 38 anadromous and freshwater wild populations (Bourret et al., 2013). The data illustrated the genetic architecture in salmon and showed the adaptive divergence of SNP allele frequencies across populations and among regional groups. Bradbury et al. applied an SNP array to Atlantic cod and showed an association between SNP allele frequencies and water temperatures across the species range (Bradbury et al., 2011). Jones et al. developed an SNP array to study geographic patterns of genetic variation on stickleback. Substantial genetic variation was found in 34 populations with predominant patterns reflecting demographic history and geographic structure. Genome regions contributing to evolution of marine–freshwater or benthic–limnetic species pairs were identified (Jones et al., 2012).

Many SNP markers have been identified from common carp (Kongchum et al., 2010; Xu et al., 2014), and a high-throughput 250 K common carp SNP array has been developed (Xu et al., 2014). The entire genome sequences of common carp have been published in 2014 (Xu et al., 2014). In the present study, genome-wide SNP genotyping was conducted to determine the genetic architecture of common carp populations and the genetic bases for environmental adaptation. A total of 2198 samples were successfully genotyped with high quality (see Materials and Methods). These samples belonged to 14 different populations, including Yellow River carp (YR), HB, Xingguo carp (XG), OUJ, KOI, Qingshuijiang carp (QSJ), JIAN, Songhe carp (SH), SP, Heilongjiang carp (HLJ), Danube carp (DANU), Szarvas 22 carp (SZ), Tisza carp (TZ), and a population from the USA (AME). YR are mainly cultured along the Yellow River basin of China; HB, XG, and OUJ are mainly cultured in the south of China; SH, SP, and HLJ were mainly cultured in the north of China, and QSJ is mainly cultured on the southwest of China. DANU, SZ, and TZ were collected from Europe, while AME was collected from Alabama in the USA. The phenotypic traits of these common carp populations differ from each other. Most common carp are black or gray, but HB and XG are red. KOI and OUJ have various skin color and patterns, including white, black, red, yellow, blue, and cream. The molecular mechanisms underlying trait differences between different common carp populations were unveiled in this study.

MATERIALS AND METHODS

Sample Collection

The 14 populations of *C. carpio* (2,198 individuals) were randomly collected across Europe, North America, and China. DANU, TZ, and Szarvas 22 (SZ) were collected from the carp live gene bank of the Research Institute for Fisheries, Aquaculture and Irrigation of Hungary (HAKI). North American carp (AME) were collected from the Chattahoochee River in Alabama in the USA. Ten other populations were sampled from China, namely, the YR from Zhengzhou of Henan Province, HB from Wuyuan of Jiangxi Province, XG from Xingguo of Jiangxi Province, OUJ

from Oujiang of Zhejiang Province, KOI from Beijing, QSJ from Guiyang of Guizhou Province, JIAN from Wuxi of Jiangsu Province, SH and SP from Harbin of Heilongjiang Province, and HLJ from Mudanjiang of Heilongjiang Province. The numbers of samples from each population are shown in **Table S1**.

DNA Extraction, Genotyping, and Quality Control

Genomic DNA was extracted from blood or fin samples using a DNeasy 96 Blood & Tissue Kit (Qiagen, Shanghai, China) following the manufacturer's protocol. Extracted DNA was quantified by a Nanodrop-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). DNA integrity was examined on a 1.0% agarose gel by electrophoresis. The final DNA concentration was diluted to 50 ng/μl for genotyping with an amount of 2 μg per sample. The common carp 250-K SNP array was developed using Affymetrix Axiom genotyping technology. Genotyping was performed by GeneSeek (Lincoln, Nebraska, USA). After genotyping, PLINK v1.9 software (<https://www.cog-genomics.org/plink2>) was used for quality control (Chang et al., 2015). SNPs with low call rate (< 95%) or low minor allele frequency (MAF < 5%) were excluded, and samples with <90% genotyping rate were filtered out. The filtered genotype file was uploaded into the European Nucleotide Archive database (<https://www.ebi.ac.uk/ena/data/view/PRJEB33066>).

Phylogeny, Principal Component Analysis, and Genetic Structure

A maximum-likelihood tree was constructed by RAxML with 1,000 bootstraps (Stamatakis, 2014), and the tree was displayed with iTOL software (<http://itol.embl.de/upload.cgi>). Principal component analysis (PCA) was conducted using GATK software (Van der Auwera et al., 2013), and all of the SNPs were used to investigate the population structure using Structure 2.3.1 software with 2,000 iterations and the MCMC model (Falush et al., 2007). The optimal *K* value was selected by Delta *K* method (Evanno et al., 2005). The resulting structure matrix was plotted using StructurePlot 2.0 software (Ramasamy et al., 2014).

LD Decay and Haplotype Construction

Linkage disequilibrium (LD) decays for the main populations and all of the samples were calculated within a range of 50 kb using PLINK (Chang et al., 2015). The average *R*² value of each 1 kb region was calculated (Zhou et al., 2016; Xu et al., 2017). All of the *R*² values were then plotted against the physical distances of SNPs in units of kb. Haplotype blocks in different populations were identified by PLINK software using the "--blocks" parameter.

Calculation of π Ratio, *F*_{st}, Tajima's *D*, and Identification of Selective Signatures

We calculated the π distribution for each linkage group using a sliding window method in Vcftools. The window width was set to 100 kb, and the stepwise distance was 100 kb. The π values from the main populations were compared, and the ratios were sorted.

*F*_{st} and Tajima's *D* values were also calculated using Vcftools with the parameters "--weir-fst-pop" and "--TajimaD," respectively. We identified the regions with the 5% highest π ratios and the regions with the 5% highest *F*_{st} values. Together with regions identified on the basis of the above two thresholds, genes within selective sweeps were annotated using GOEAST for Gene Ontology (GO) (Young et al., 2010) and the DAVID software (Dennis et al., 2003) for KEGG pathway analysis. Scatter plots of π and *F*_{st} values were generated using the ggplot2 package of the Comprehensive R Archive Network (<http://cran.r-project.org/package=ggplot2>).

RESULTS AND DISCUSSION

Sample Collection and Genotyping

A total of 14 populations of *C. carpio* were collected from 13 locations in China, Hungary, and the USA (**Figure 1**). In addition to the geographic divergences of these populations, several populations were included due to their special biological features, such as scale pattern (SP population), red body color (HB population, XG population, and QH group in the OUI population), and purse-like body shape (HB population). After DNA extraction and SNP genotyping using Carp 250-K SNP array, a raw genotype database of 222,694 SNPs for 2,198 samples was generated. A total of 2,198 samples with 134,719 polymorphic SNPs passed the quality control threshold and were used for further analysis. Sample information for each population is shown in **Table S1**.

Phylogeny, PCA, and Population Structure

To investigate the divergence of the representative *C. carpio* populations from different locations, we constructed the phylogenetic tree using the whole genotyping data (**Figure 2A**). The Asian and European populations formed two distinct clades, while the AME population grouped with the European clade. The SP population constituted the major part of the European clade as it was bred from mirror carp originally introduced from Europe in the 1950s. A similar result was indicated by PCA, showing subgroups in either Asian populations or European populations (**Figure 2B**). In the Asian populations, YR formed a tight cluster and other populations (JIAN, HB, XG, KOI, OUI, and QSJ) formed another subgroup. Three subgroups could be identified in the European cluster. The first subgroup contained only SP samples which were closely grouped. The second subgroup included three Hungary populations (DANU, TZ, and SZ), SP population, and HLJ population, indicating close relationship among SP, HLJ, and Hungary carp. A small number of SP samples and all of the AME samples formed the third subgroup, showing that the common carp in the USA might have originated from European populations. We analyzed the population structure using the Bayesian clustering program STRUCTURE. Since the values of Delta *K* from ln likelihood were high for the models *K* = 5, we showed the clusters of *K* = 5 in **Figure 2C**. The Asian populations were separated into two subgroups, similar to the PCA result, and the European populations showed shared common ancestry. Within the Asian populations, obvious genetic admixture was observed in the YR population and other populations, except

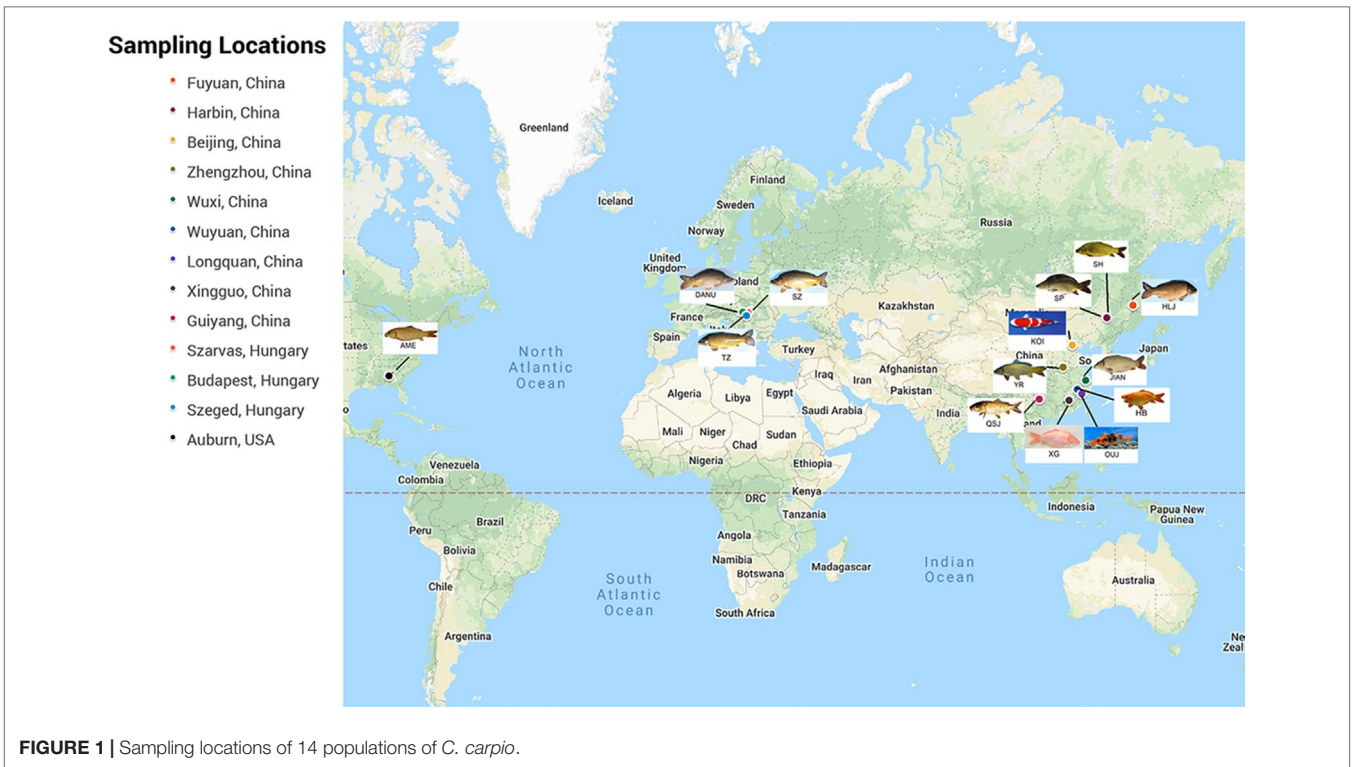


FIGURE 1 | Sampling locations of 14 populations of *C. carpio*.

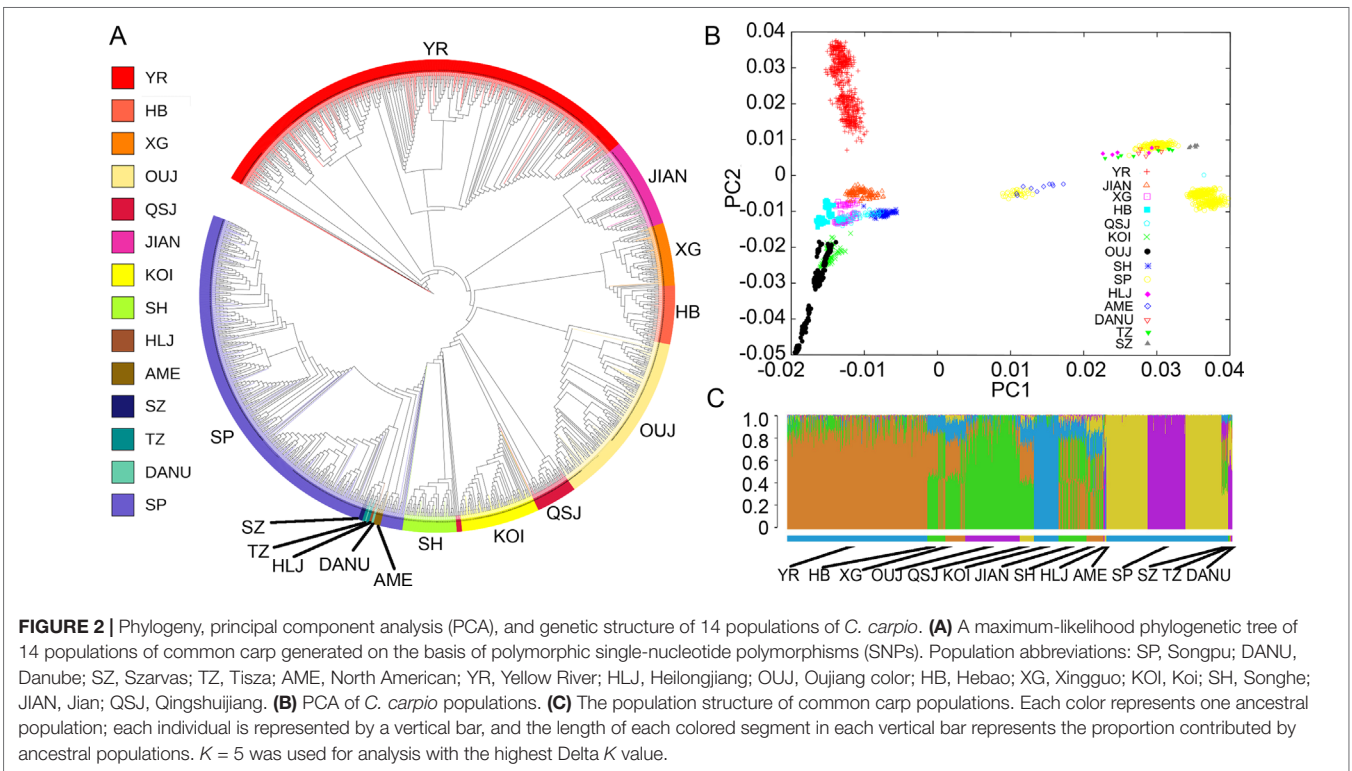


FIGURE 2 | Phylogeny, principal component analysis (PCA), and genetic structure of 14 populations of *C. carpio*. **(A)** A maximum-likelihood phylogenetic tree of 14 populations of common carp generated on the basis of polymorphic single-nucleotide polymorphisms (SNPs). Population abbreviations: SP, Songpu; DANU, Danube; SZ, Szarvas; TZ, Tisza; AME, North American; YR, Yellow River; HLJ, Heilongjiang; OUJ, Oujiang color; HB, Hebao; XG, Xingguo; KOI, Koi; SH, Songhe; JIAN, Jian; QSJ, Qingshuijiang. **(B)** PCA of *C. carpio* populations. **(C)** The population structure of common carp populations. Each color represents one ancestral population; each individual is represented by a vertical bar, and the length of each colored segment in each vertical bar represents the proportion contributed by ancestral populations. $K = 5$ was used for analysis with the highest Delta K value.

for KOI and OUI. The KOI population has had a highly inbred history to maintain the purity of the genetic component, and the OUI population also showed relatively pure genetic structure due to their habitats in the isolated mountainous areas in Zhejiang, China. This result is in accordance with our previous study (Xu et al., 2014), but there were slight differences in the shared genetic components in all of the populations.

was consistent with the population structure results. Haplotypes were constructed for all samples and seven populations using PLINK. The distribution of different haplotype lengths was calculated by ggplot2 package, and OUI, SP, XG, and YR showed longer blocks (Figure 3B and Table S2). Haplotypes are useful in GWAS analysis and provide more SNP genotyping information through imputation.

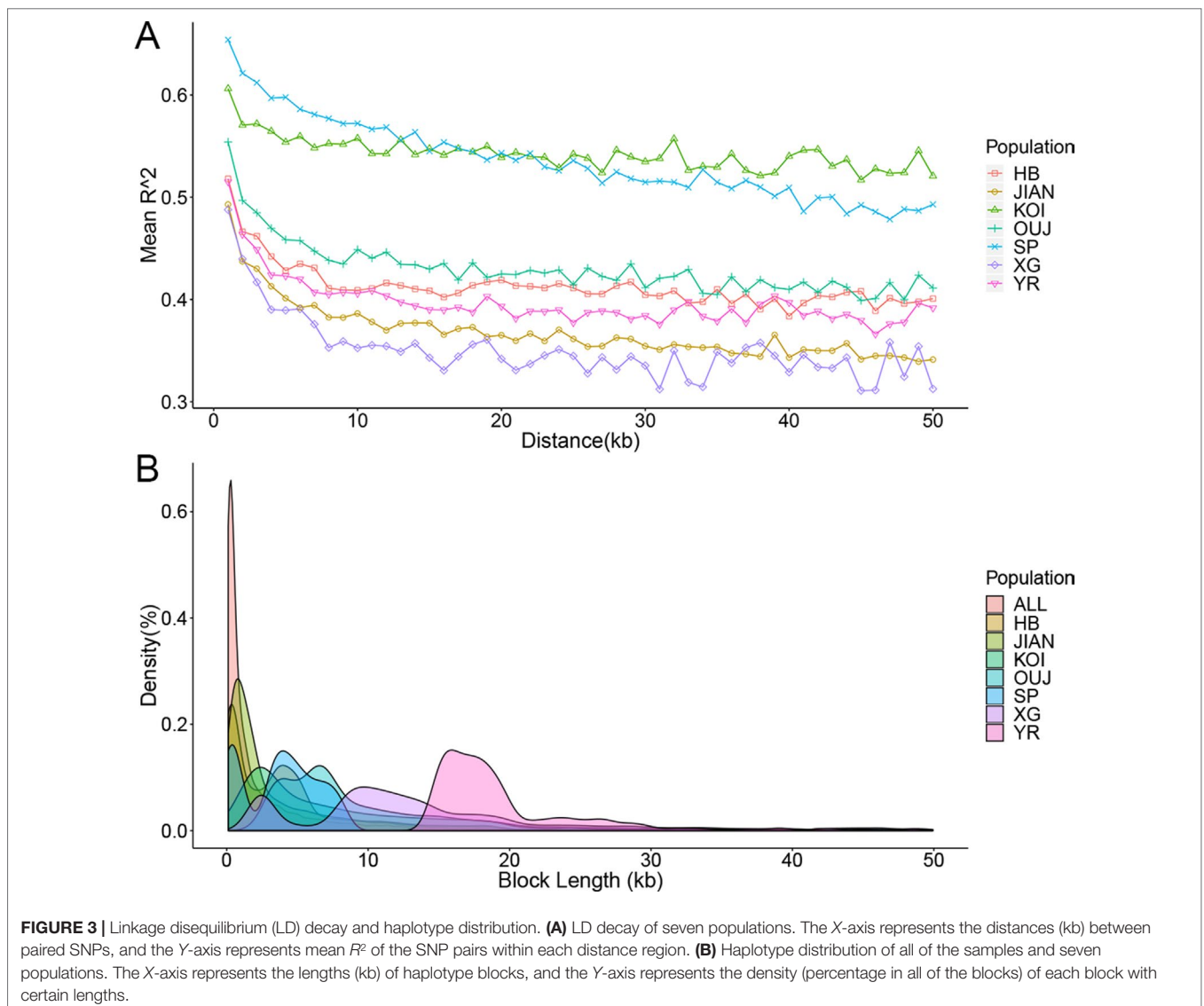
LD Decay and Haplotype Construction

Based on the genome assembly and SNP array for *C. carpio*, the LD was investigated for seven main populations that contained an adequate number of samples. The R^2 value among each pair of SNPs was calculated using PLINK software, and the raw data were classified by distance ranges. The LD decays with the extension of the distance between SNPs and different populations showed distinct LD decay patterns (Figure 3A). KOI and OUI had significantly higher R^2 values than five other populations, which

Genome-wide Selective Sweep Analysis

C. carpio is a genetically diverse species that has adapted to a variety of environments in Eurasia and has been domesticated for more than 2,000 years (Xu et al., 2014). *C. carpio* has been bred into numerous strains, generating distinct phenotypes in body color, scale pattern, and body shape. These characters are partially attributable to genome diversity due to environmental adaptation.

The genetic diversity in certain genome regions might be reduced due to natural selection. To identify the genome regions



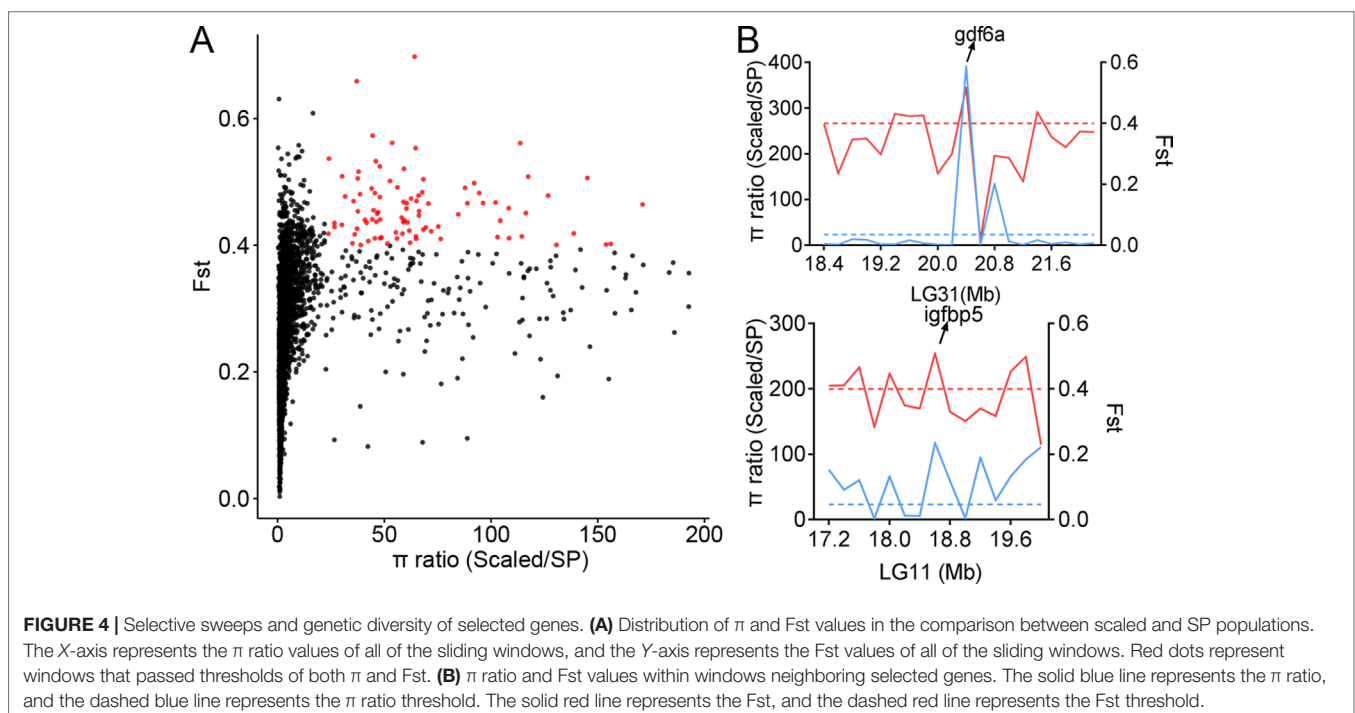
under selective pressure in populations with distinct biological features, we scanned the genome-wide variations and allele frequency spectra of the 134,719 SNPs. The areas of comparison included scales, body shape, and body color. The π ratios of three groups (π Scaled carp/SP, π YR/HB, and π QH/FY) were calculated using a 200-kb sliding-window approach with Vcftools software. In comparison to the Scaled carp in Asia (including YR, HB, XG, OUI, QSI, KOI, and JIAN), the SP population had distinct genetic diversities across the whole genome (Figure 4A). We identified 321 significant windows corresponding to 64.2 Mb in size (top 5%, empirical π ratios ≥ 23.17), which included 289 candidate genes based on the π ratio analysis. To validate the genome regions under strong selective sweeps in the SP population, the genome regions with F_{st} greater than 0.3995 (top 5%) were also identified, corresponding to 64.2 Mb and 285 candidate genes. A total of 100 candidate genes shared by both the π ratio and F_{st} analysis were identified as potentially affected genes under selective sweeps (Figure 4A, Table S3).

The results suggest that the genomes of the SP population have been significantly altered, by the environment, into a no-scale pattern. Fish scale is an epidermal appendage and it is an important protective tissue. Among the samples, all were fully scaled except for the SP carp that lacks scales. Through selection sweeping analysis of comparing SP carp with other fully scaled carps, the significant genome regions were identified (Table S3). Also, several target genes were found that might be related to the scale pattern (Figure 4B). Growth differentiation factor 6a (*gdf6a*), also named cartilage derived morphogenetic protein 2, is a member of the BMP family. The expression of the *gdf6a* gene has been detected in both fetal and post-natal cartilaginous tissues involved in the development of long bones. The *gdf6a* gene may play a role in suppression of ossification (Wei et al., 2016). Insulin-like growth factor-binding

protein 5 (*igfbp5*) is a member of the IGFBP family that can either inhibit or stimulate the growth-promoting effects of the IGFs on cell culture. It is involved in the Ras/p38 MAPK signaling pathway in regulating cell proliferation and apoptosis (Yang et al., 2018). To study these candidate genes and their potential functions, GO and KEGG analyses were performed on the candidate genes, offering insight into the genetic evolution and adaptive mechanisms of the SP population (Tables S4 and S9).

We also investigated selective sweeps in comparisons (YR vs HB, QH vs FY), which showed distinct body shapes or body color, respectively. YR and HB were compared because they represent the typical populations in central China and southern China, respectively. QH and FY were compared due to their high similarity in genome background and the possibility that it might be easier to screen out potential genes associated with body color. We identified a total of 321 significant windows corresponding to 64.2 Mb in size for each comparison (top 5%, empirical π ratios ≥ 1.3121 , 1.9341, respectively), which included 293 and 287 candidate genes, respectively. To further validate the genome regions under strong selective sweeps in the HB or FY population, the genome regions with F_{st} greater than 0.2588 or 0.1622 (top 5%) were also identified, including 278 and 292 candidate genes, respectively. A total of 38 and 65 candidate genes shared by both the π ratio and F_{st} analysis were recognized, respectively, as genes potentially affected under selective sweeps (Figures S1A and S2A, Tables S5 and S7).

The purse-like shape of the HB population was probably due to extensive growth of muscle or bones compared to the YR population, and several genes (*trhr* and *bmpr1b*) were screened out for their potential functions in bone and muscle development (Figure S1B). Previous genome-wide association and replication studies identified *trhr* as a gene associated with lean body mass (Liu et al., 2009). The gene *bmpr1b* is engaged in the regulation



of skeletal development through interactions with FGFR families (Qi et al., 2014). GO enrichment analysis was also performed on the candidate genes, offering insight into the genetic evolution and adaptive mechanisms of the HB population (Tables S6 and S9). GO terms including thyrotropin-releasing hormone receptor activity, Wnt-activated receptor activity, and transforming growth factor beta-activated receptor activity were enriched, providing clues for more detailed analysis. Coloration is an important trait for common carp, especially for ornamental strains, since it is often a criterion for visually determining quality and market value. Ouj carp is a famous ornamental farmed fish, which has four distinct color patterns, namely, whole white (FY), whole red (QH), white with scattered big black spots, and red with scattered big black spots. FY and QH were compared, and several potential target genes were identified (Figure S2B). The keratinocyte growth factor (*fgf7/kgf*) can promote melanosome transfer and act on recipient keratinocytes through stimulation of the phagocytic process. *Fgf7* affects keratinocytes derived from different skin color (Cardinali et al., 2008). Another gene in selective sweep regions, *opsin5*, has been reported engaging in phototransduction and regulates seasonal changes in color perception (Shimmura et al., 2017). It was identified as an ultraviolet (UV)-sensitive pigment of the retina and other photosensitive organs in birds (Ohuchi et al., 2012). GO and KEGG analyses were also performed on the candidate genes and offered insight into the genetic evolution and adaptive mechanisms of the SP population (Tables S8 and S9). Several significant pathways were enriched, including dopamine receptor signaling pathway and dopamine neurotransmitter receptor activity. This indicated the importance of dopamine-related networks in body color determination. The representative GO terms and pathways enriched in these comparisons were selected, and the *Fst* values of genes in these pathways were significantly higher than the whole-genome level (Figure S3).

CONCLUSIONS

We investigated the genomic divergence among various populations of *C. carpio*. Distinct genetic component differences were identified between Asian and European populations. The haplotypes of each population could benefit research on trait associations. Selective sweep analyses results showed that hundreds of genes within selective sweep regions were identified by genome scanning among different populations, including *gdf6a*, *bmpr1b*, and *opsin5*. This study comprehensively revealed genetic structure of global populations of *C. carpio*, and potential trait-related genes could be valuable for genome-assisted breeding of *C. carpio*.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the care and use of animals for scientific purposes set up by the Animal Care and Use Committee of Chinese Academy of Fishery Sciences (ACUC-CAFS). The protocol was approved by the ACUC-CAFS. Before the blood or fin samples were collected, all of the fishes were euthanized in MS222 solution.

AUTHOR CONTRIBUTIONS

PX initiated and coordinated the research project. JX and YJ conducted the analysis and drafted the manuscript. ZZ, HZ, and JF engaged in sample collection and genotyping analysis. CD, WP, BC, and RT took part in enrichment analysis. All authors read and approved the final manuscript.

FUNDING

This work was supported by Central Public-Interest Scientific Institution Basal Research Fund, CAFS (No. 2016GH02, No. 2016HY-JC0301, and No. 2016HY-ZD0302), the National Natural Science Foundation of China (No. 31502151 and No. 31422057), the National High-Technology Research and Development Program of China (2011AA100401), the National Key Research and Development Program (2018YFD0900102), and the National Infrastructure of Fishery Germplasm Resources of China (No. 2018DKA30470).

ACKNOWLEDGMENTS

We thank Mr. Xianhu Zheng, Mr. Youyi Kuang, Mr. Xiaowen Sun, Mr. Chenghui Wang, and Mr. Zhanjiang Liu for their assistance with sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00660/full#supplementary-material>

SUPPLEMENTARY TABLES 1–9 | Summary of genotyping data, haplotype blocks, and selective sweep results.

SUPPLEMENTARY FIGURE 1 | π , *Fst* distribution, and selected genes in selective sweep analysis between YR and HB populations. (A) Distribution of π and *Fst* values in the comparison between YR and HB populations. The X-axis represents the π ratio values of all sliding windows, and the Y-axis represents the *Fst* values of all sliding-windows. Red dots represent windows that passed thresholds of both π and *Fst*. (B) π ratio and *Fst* values within windows neighboring selected genes. The solid blue line represents the π ratio, and the dashed blue line represents the π ratio threshold. The solid red line represents the *Fst*, and the dashed red line represents the *Fst* threshold.

SUPPLEMENTARY FIGURE 2 | π , *Fst* distribution, and selected genes in selective sweep analysis between QH and FY samples. (A) Distribution of π and *Fst* values in the comparison between QH and FY samples. The X-axis represents the π ratio values of all sliding windows, and the Y-axis represents the *Fst* values of all sliding windows. Red dots represent windows that passed thresholds of both π and *Fst*. (B) π ratio and *Fst* values within windows neighboring selected genes. The solid blue line represents the π ratio, and the dashed blue line represents the π ratio threshold. The solid red line represents the *Fst*, and the dashed red line represents the *Fst* threshold.

SUPPLEMENTARY FIGURE 3 | Box plot of the *Fst* differences between representative GO terms and the whole genome. Boxes denote the values between the 25th and 75th percentiles and the black transverse line inside the box denotes the median. Black vertical lines denote the values within 1.5 times of quartile values. Outliers are shown as black dots. (A) Scaled vs SP. (B) YR vs HB. (C) QH vs FY.

REFERENCES

- Balon, E. K. (1995). Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers. *Aquaculture* 129, 46. doi: 10.1016/0044-8486(94)00227-F
- Bartfai, R., Egedi, S., Yue, G. H., Kovacs, B., Urbanyi, B., Tamas, G., et al. (2003). Genetic analysis of two common carp broodstocks by RAPD and microsatellite markers. *Aquaculture* 219, 11. doi: 10.1016/S0044-8486(02)00571-9
- Bostock, J., McAndrew, B., Richards, R., Jauncey, K., Telfer, T., Lorenzen, K., et al. (2010). Aquaculture: global status and trends. *Phil. Trans. R. Soc. B* 365, 16. doi: 10.1098/rstb.2010.0170
- Bourret, V., Kent, M. P., Primmer, C. R., Vasemagi, A., Karlsson, S., Hindar, K., et al. (2013). SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Mol. Ecol.* 22 (3), 532–551. doi: 10.1111/mec.12003
- Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Paterson, I. G., Snelgrove, P. V., et al. (2011). Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. *Mol. Ecol. Resour.* 11 Suppl 1, 218–225. doi: 10.1111/j.1755-0998.2010.02949.x
- Cardinali, G., Bolasco, G., Aspite, N., Lucania, G., Lotti, L. V., Torrisi, M. R., et al. (2008). Melanosome transfer promoted by keratinocyte growth factor in light and dark skin-derived keratinocytes. *J. Invest. Dermatol.* 128 (3), 558–567. doi: 10.1038/sj.jid.5701063
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi: 10.1186/s13742-015-0047-8
- Cheng, L., Liu, L., Yu, X., Wang, D., and Tong, J. (2010). A linkage map of common carp (*Cyprinus carpio*) based on AFLP and microsatellite markers. *Anim. Genet.* 41 (2), 191–198. doi: 10.1111/j.1365-2052.2009.01985.x
- Chistiakov, D. A., and Voronova, N. V. (2009). Genetic evolution and diversity of common carp *Cyprinus carpio* L. *Cent. Eur. J. Biol.* 4, 9. doi: 10.2478/s11535-009-0024-2
- Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4 (5), P3. doi: 10.1186/gb-2003-4-9-r60
- Dong, C., Xu, J., Wang, B., Feng, J., Jeney, Z., Sun, X., et al. (2015). Phylogeny and evolution of multiple common carp (*Cyprinus carpio* L.) populations clarified by phylogenetic analysis based on complete mitochondrial genomes. *Mar. Biotechnol. (N.Y.)* 17 (5), 565–575. doi: 10.1007/s10126-015-9639-7
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14 (8), 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7 (4), 574–578. doi: 10.1111/j.1471-8286.2007.01758.x
- FAO Fisheries and Aquaculture Department. (2007). *The State of World Fisheries and Aquaculture 2006*. Rome: Food and Agriculture Organization of the United Nations.
- Froufe, E., Magyary, I., Lehoczy, I., and Weiss, S. (2002). mtDNA sequence data supports an Asian ancestry and single introduction of the common carp into the Danube Basin. *J. Fish Biol.* 61, 4. doi: 10.1111/j.1095-8649.2002.tb01756.x
- Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., et al. (2012). A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* 22 (1), 83–90. doi: 10.1016/j.cub.2011.11.045
- Kohlmann, K., and Kersten, P. (2013). Deeper insight into the origin and spread of European common carp (*Cyprinus carpio carpio*) based on mitochondrial D-loop sequence polymorphisms. *Aquaculture* 376, 8. doi: 10.1016/j.aquaculture.2012.11.046
- Kongchum, P., Palti, Y., Hallerman, E. M., Hulata, G., and David, L. (2010). SNP discovery and development of genetic markers for mapping innate immune response genes in common carp (*Cyprinus carpio*). *Fish Shellfish Immunol.* 29 (2), 356–361. doi: 10.1016/j.fsi.2010.04.013
- Liu, X., Ye, X., Liang, H., Zhou, L., Zhou, X., Zou, G., et al. (2019). Mitochondrial genome sequences reveal the evolutionary relationship among different common carp varieties (*Cyprinus carpio* L.). *Meta Gene* 19, 82–90. doi: 10.1016/j.mgene.2018.11.001
- Liu, X. G., Tan, L. J., Lei, S. F., Liu, Y. J., Shen, H., Wang, L., et al. (2009). Genome-wide association and replication studies identified TRHR as an important gene for lean body mass. *Am. J. Hum. Genet.* 84 (3), 418–423. doi: 10.1016/j.ajhg.2009.02.004
- Mabuchi, K., Senou, H., and Nishida, M. (2008). Mitochondrial DNA analysis reveals cryptic large-scale invasion of non-native genotypes of common carp (*Cyprinus carpio*) in Japan. *Mol. Ecol.* 17 (3), 796–809. doi: 10.1111/j.1365-294X.2007.03626.x
- Ohuchi, H., Yamashita, T., Tomonari, S., Fujita-Yanagibayashi, S., Sakai, K., Noji, S., et al. (2012). A non-mammalian type opsin 5 functions dually in the photoreceptive and non-photoreceptive organs of birds. *PLoS One* 7 (2), e31534. doi: 10.1371/journal.pone.0031534
- Perez-Enriquez, R., Robledo, D., Houston, R. D., and Llera-Herrera, R. (2018). SNP markers for the genetic characterization of Mexican shrimp broodstocks. *Genomics* 110 (6), 423–429. doi: 10.1016/j.ygeno.2018.10.001
- Qi, H. B., Jin, M., Duan, Y. Q., Du, X. L., Zhang, Y. Q., Ren, F. L., et al. (2014). FGFR3 induces degradation of BMP type I receptor to regulate skeletal development. *BBA-Mol. Cell Res.* 1843 (7), 1237–1247. doi: 10.1016/j.bbamcr.2014.03.011
- Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., and Naik, V. G. (2014). Structure plot: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus* 3, 431. doi: 10.1186/2193-1801-3-431
- Shimmura, T., Nakayama, T., Shinomiya, A., Fukamachi, S., Yasugi, M., Watanabe, E., et al. (2017). Dynamic plasticity in phototransduction regulates seasonal changes in color perception. *Nat. Commun.* 8 (1), 412. doi: 10.1038/s41467-017-00432-8
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9), 1312–1313. doi: 10.1093/bioinformatics/btu033
- Torati, L. S., Taggart, J. B., Varela, E. S., Araripe, J., Wehner, S., and Migaud, H. (2019). Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genet.* 20 (1), 13. doi: 10.1186/s12863-018-0711-y
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43 (1), 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43
- Wang, C., Liu, H., Lou, Z., Wang, J., Zou, J., and Li, C. (2010). Mitochondrial genetic diversity and gene flow of common carp from main river drainages in China. *Freshwater Biol.* 55, 11. doi: 10.1111/j.1365-2427.2010.02424.x
- Wang, C. H., and Li, S. F. (2004a). Genetic variability and relationships in mitochondrial DNA CO II gene sequence of red common carps in China. *Acta Genet. Sin.* 31 (11), 1226–1231.
- Wang, C. H., and Li, S. F. (2004b). Phylogenetic relationships of ornamental (koi) carp, Oujiang color carp and long-fin carp revealed by mitochondrial DNA COII gene sequences and RAPD analysis. *Aquaculture* 231, 9. doi: 10.1016/j.aquaculture.2003.10.040
- Wei, A. Q., Shen, B. J., Williams, L. A., Bhargava, D., Gulati, T., Fang, Z. M., et al. (2016). Expression of growth differentiation factor 6 in the human developing fetal spine retreats from vertebral ossifying regions and is restricted to cartilaginous tissues. *J. Orthopaedic Res.* 34 (2), 279–289. doi: 10.1002/jor.22983
- Xu, J., Li, J. T., Jiang, Y. L., Peng, W. Z., Yao, Z. L., Chen, B. H., et al. (2017). Genomic basis of adaptive evolution: the survival of amur ide (*Leuciscus waleckii*) in an extremely alkaline environment. *Mol. Biol. Evol.* 34 (1), 145–159. doi: 10.1093/molbev/msw230
- Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., et al. (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat. Genet.* 46 (11), 1212–1219. doi: 10.1038/ng.3098
- Yang, Z. G., Wang, J. F., Pan, Z. Z., and Zhang, Y. H. (2018). miR-143-3p regulates cell proliferation and apoptosis by targeting IGF1R and IGFBP5 and regulating the Ras/p38 MAPK signaling pathway in rheumatoid arthritis. *Exp. Ther. Med.* 15 (4), 3781–3790. doi: 10.3892/etm.2018.5907

- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11 (2), R14. doi: 10.1186/gb-2010-11-2-r14
- Zhou, J., Wang, Z., Ye, Y., and Wu, Q. (2003). PCR-RFLP analysis of mitochondrial DNA ND5/6 region among 3 subspecies of common carp (*Cyprinus carpio* L.) and its application to genetic discrimination of subspecies Chinese. *Sci. Bull.* 48, 4. doi: 10.1360/03tb9098
- Zhou, J., Wu, Q., Wang, Z., and Ye, Y. (2004). Genetic variation analysis within and among six varieties of common carp (*Cyprinus carpio* L.) in China using microsatellite markers. *Genetika* 40 (10), 1389–1393. doi: 10.1023/B:RUGE.0000044758.51875.25
- Zhou, L. Y., Liu, S. Y., Wu, W. X., Chen, D. B., Zhan, X. D., Zhu, A. K., et al. (2016). Dissection of genetic architecture of rice plant height and heading date by multiple-strategy-based association studies. *Sci. Rep.* 6, 29718. doi: 10.1038/srep29718

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor is currently editing co-organizing a Research Topic with one of the authors PX, and confirms the absence of any other collaboration.

Copyright © 2019 Xu, Jiang, Zhao, Zhang, Peng, Feng, Dong, Chen, Tai and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.