Wu et al. *Zool. Res.* 2024, 45(2): 329−340
https://doi.org/10.24272/j.issn.2095-8137.2023.270

ZR | Zoological Research

Article Open Access

Deciphering the population structure and genetic basis of growth traits from whole-genome resequencing of the leopard coral grouper (*Plectropomus leopardus*)

Shao-Xuan Wu[1,#], Qi-Fan Zeng[1,2,#], Wen-Tao Han[1,#], Meng-Ya Wang[1], Hui Ding[1], Ming-Xuan Teng[1], Ming-Yi Wang[1], Pei-Yu Li[1], Xin Gao[1], Zhen-Min Bao[1,2,3], Bo Wang[1,2,*], Jing-Jie Hu[1,2,3,*]

[1] *MOE Key Laboratory of Marine Genetics and Breeding*, *College of Marine Life Sciences/Key Laboratory of Tropical Aquatic Germplasm of Hainan Province*, *Sanya Oceanographic Institution*, *Ocean University of China*, *Qingdao*, *Shandong/Sanya*, *Hainan* 266100/572025, *China*

[2] *Hainan Seed Industry Laboratory*, *Sanya*, *Hainan* 572025, *China*

[3] *Southern Marine Science and Engineer Guangdong Laboratory*, *Guangzhou*, *Guangdong* 511458, *China*

ABSTRACT

The leopard coral grouper (*Plectropomus leopardus*) is a species of significant economic importance. Although artificial cultivation of *P. leopardus* has thrived in recent decades, the advancement of selective breeding has been hindered by the lack of comprehensive population genomic data. In this study, we identified over 8.73 million single nucleotide polymorphisms (SNPs) through whole-genome resequencing of 326 individuals spanning six distinct groups. Furthermore, we categorized 226 individuals with high-coverage sequencing depth (≥14×) into eight clusters based on their genetic profiles and phylogenetic relationships. Notably, four of these clusters exhibited pronounced genetic differentiation compared with the other populations. To identify potentially advantageous loci for *P. leopardus*, we examined genomic regions exhibiting selective sweeps by analyzing the nucleotide diversity ($\theta\pi$) and fixation index (F_{ST}) in these four clusters. Using these high-coverage resequencing data, we successfully constructed the first haplotype reference panel specific to *P. leopardus*. This achievement holds promise for enabling high-quality, cost-effective imputation methods. Additionally, we combined low-coverage sequencing data with imputation techniques for a genome-wide association study, aiming to identify candidate SNP loci and genes associated with growth traits. A significant concentration of these genes was observed on chromosome 17, which is primarily involved in skeletal muscle and embryonic development and cell proliferation. Notably, our detailed investigation of growth-related SNPs across the eight clusters revealed that cluster 5 harbored the most promising candidate SNPs, showing potential for genetic selective breeding efforts. These findings provide a robust toolkit and valuable insights into the management of germplasm resources and genome-driven breeding initiatives targeting *P. leopardus*.

Keywords: *Plectropomus leopardus*; Whole-genome resequencing; Growth; Haplotype reference panel; Single nucleotide polymorphisms

INTRODUCTION

Groupers, belonging to the family Epinephelidae in the order Perciformes encompass 165 species, spanning 16 genera. These remarkable marine fish play a pivotal role as apex predators in coral reef ecosystems (Zhuang et al., 2013), primarily inhabiting coastal regions in tropical and subtropical zones. At present, 47 distinct grouper species have been actively cultivated in East and Southeast Asia (Rimmer & Glamuzina, 2019), with China accounting for approximately 82% of the total production (Rimmer & Glamuzina, 2019).

The leopard coral grouper (*Plectropomus leopardus*) has gained prominence among grouper species. This species inhabits the Indo-Pacific Ocean, extending from southern Japan to Australia and eastward to the Caroline Islands, Fiji, and Tonga (Froese & Pauly, 2023). It is noted for its crimson

Received: 25 August 2023; Accepted: 10 October 2023; Online: 06 December 2023

Foundation items: This study was supported by the National Key Research and Development Program of China (2022YFD2400501), Key R&D Project of Hainan Province (ZDYF2021XDNY133), Project of Sanya Yazhouwan Science and Technology City Management Foundation (SKJC-2020-02-009), PhD Scientific Research and Innovation Foundation of Sanya Yazhou Bay Science and Technology City (HSPHDSRF-2022-02-007) and Young Elite Scientists Sponsorship Program by CAST (2023QNRC001)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: wb@ouc.edu.cn; hujingjie@ouc.edu.cn

hue and delicious, low-fat, high-protein muscle. The resulting market demand has led to a precipitous decline in its wild populations (Yang et al., 2020a). Although establishing a breeding industry for *P. leopardus* has become critical, such an endeavor hinges on the application of cutting-edge molecular biology and genomics technology.

Scientists have undertaken extensive molecular biology research to improve the productivity and economic viability of *P. leopardus*, encompassing areas such as nutrition (Mekuchi et al., 2018), body color modulation (Liu et al., 2022), and rearing condition optimization (Xia et al., 2020). Furthermore, advancements in genomic studies have been instrumental in the development of microsatellite markers (Zhang et al., 2010) and in the establishment of multiple chromosome-level genomes of *P. leopardus* (Han et al., 2023; Wang et al., 2020; Zhou et al., 2020). Despite these developments, however, there remains a crucial gap in our understanding of germplasm resources, particularly regarding population structure and genetic diversity.

Although diverse markers have been employed in aquaculture population genetics, including simple sequence repeat markers (Kessuwan et al., 2016), inter-simple sequence repeat markers (Bielikova et al., 2021), and random amplified polymorphic DNA markers (Parvez et al., 2021), they suffer from limited detection efficiencies. The advent of reference genomes has made whole-genome resequencing (WGRS) a potent genotyping approach for elucidating population structures and genetic variations (Rubinacci et al., 2021), and has been widely applied across various economically important species (Wang et al., 2022). Hence, a genome-wide exploration of genetic diversity and population differentiation using WGRS data is essential for advancing genetic breeding efforts in *P. leopardus*.

The emergence of next-generation sequencing technology has made WGRS a key tool in identifying marker loci associated with phenotypic traits. Genome-wide association studies (GWAS) are particularly efficient in identifying genotype-phenotype associations (Tam et al., 2019). In groupers, GWAS have been conducted across a range of quantitative and qualitative traits, primarily using low-coverage WGRS data (Wu et al., 2019; Yang et al., 2020b; Yu et al., 2018). However, low-coverage WGRS may result in high rates of missing genotypic data. Addressing this issue requires the integration of a reliable haplotype reference panel, which not only improves genotype accuracy but also reduces sequencing costs (Marchini & Howie, 2010). The success of imputation techniques has been demonstrated across various genotyping platforms in aquatic species (Zeng et al., 2022). Imputing missing data enhances the precision of identifying and utilizing genetic variants of varying frequencies within a population, thereby facilitating the screening of key loci and favorable alleles linked to significant phenotypes.

Considering its growing significance in aquaculture, the management of germplasm resources and development of breeding initiatives for *P. leopardus* have become increasingly important, coinciding with the continuous expansion of cultivation efforts. In this comprehensive study, we used WGRS data from 326 individuals representing six distinct groups of *P. leopardus* in Hainan Province, China. Utilizing this dataset, we performed in-depth analyses of population structure and genetic effects using single nucleotide polymorphism (SNP) datasets. Our extensive data facilitated the creation of the first haplotype reference panel specific to *P. leopardus*, providing high-accuracy imputation capabilities for genetic analyses. Additionally, we applied low-coverage WGRS combined with high-accuracy imputation techniques for GWAS. Our results revealed key SNP loci and advantageous alleles linked to growth traits in *P. leopardus*, relevant for future research and practical breeding applications. Based on these findings, we explored the genetic diversity within cultured *P. leopardus* populations, offering valuable insights for enhancing germplasm management and streamlining genetic improvements in selective breeding programs (Wang et al., 2023).

MATERIALS AND METHODS

Sample collection and genome resequencing

A total of 272 leopard coral groupers were collected from five distinct cultured groups on Hainan Island, China: Wenchang (WC, 21 individuals, 10 years post-hatching), Dongfang (DF, 159 individuals, 6 months post-hatching), Wanning (WN, 17 individuals, 2 years post-hatching), Lingshui (LS, 39 individuals, 4 years post-hatching), and Sanya (SY, 36 individuals, 1 year post-hatching) (Figure 1A). Each of these cultured groups originated from the crossbreeding of different parental populations obtained from various sources, resulting in a remarkably complex genetic lineage.

To facilitate genetic analyses, we sampled caudal fin tissues, which were then preserved in 95% absolute ethanol at −20°C. Genomic DNA was extracted from each sample (200 mg) using a Marine Animal Tissue Genomic DNA Extraction Kit (Tiangen Biotech, China). Subsequently, the quality and quantity of the genomic DNA were assessed using a Qubit 4 fluorometer (Thermo Fisher Scientific, USA). Paired-end libraries with a 300 bp insert size were prepared using a VAHTS Universal DNA Library Prep Kit for Illumina V3 (Vazyme, China), following standard protocols, with sequencing performed using the Illumina NovaSeq 6000 platform (USA).

To construct comprehensive haplotype profiles for *P. leopardus*, a subgroup of 172 individuals selected from the five sampling sites was subjected to high-coverage sequencing, yielding 17 Gb of reads per sample. The remaining 100 individuals were subjected to low-coverage sequencing, generating 5 Gb of reads per sample. The dataset was further enriched by the inclusion of a WGRS dataset collected by the Yellow Sea Fisheries Research Institute (YS), which included 54 individuals, each sequenced at a depth of approximately 14× (Zhou et al., 2020). Overall, our comprehensive WGRS dataset included data from a cohort of 326 individuals.

Population genetic structure and evolutionary analysis

High-coverage resequencing data (minimum depth of ≥14×) from a cohort of 226 individuals spanning six distinct groups were used to investigate the population genetic structure and evolutionary dynamics of *P. leopardus*. Population genetics of *P. leopardus* were assessed using the pairwise fixation index (F_{ST}), ancestral estimation, phylogenetic analysis, and other well-established methodologies for analyzing population structure (Meirmans & Hedrick, 2011; Peter, 2016). Ancestral estimation was performed using ADMIXTURE v1.3.0 (Alexander et al., 2009), assessing a range of *K*-values (representing putative number of genetic groups), with consideration of the lowest cross-validation error. To further

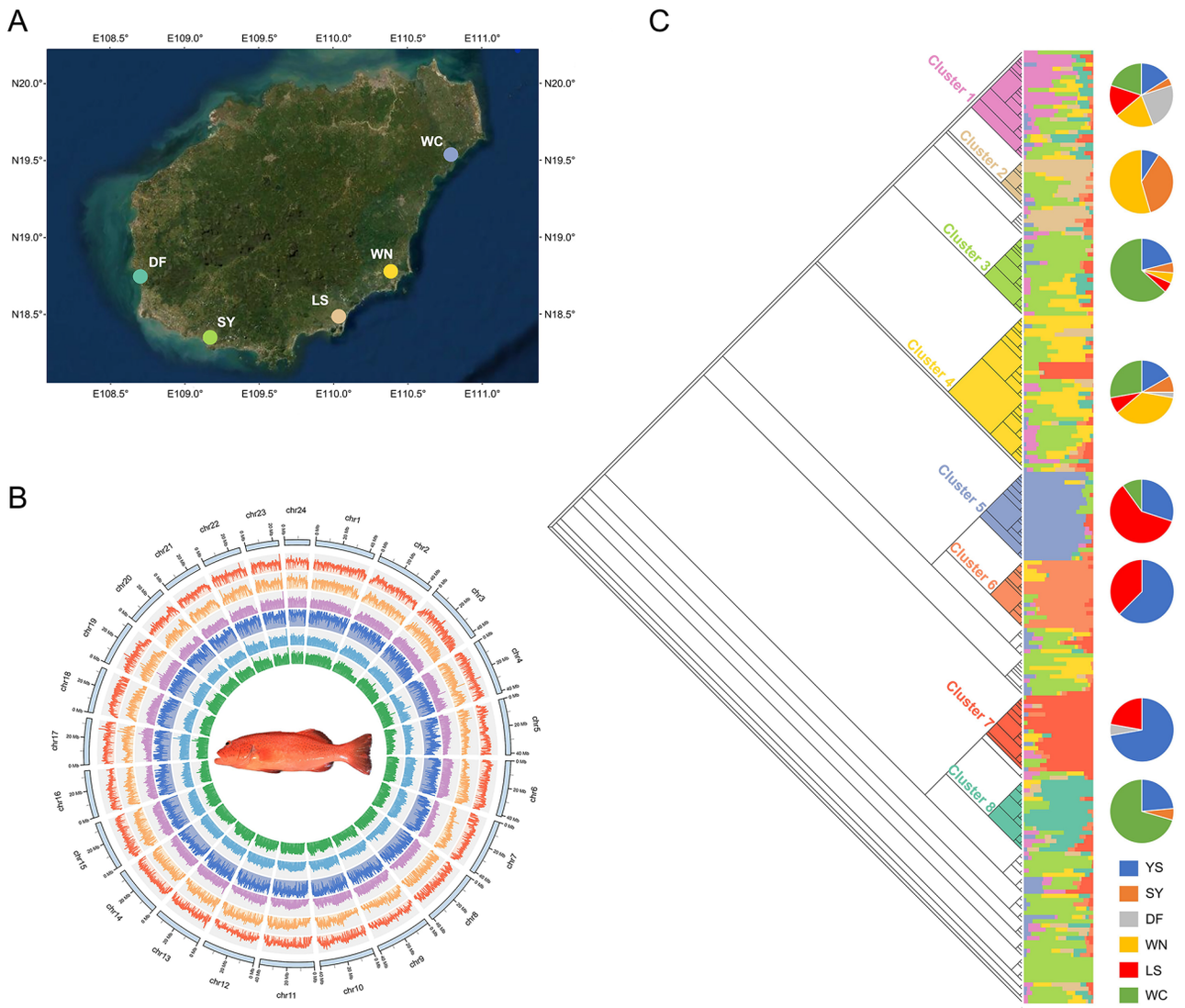

Figure 1 SNP characterization of cultured groups and preliminary results of population structure analysis

A: Sampling location of five *P. leopardus* groups: Wenchang (WC), Wanning (WN), Lingshui (LS), Sanya (SY), Dongfang (DF). B: Density of SNP loci in 24 chromosomes of six groups, WC, YS, DF, WN, LS, and SY from outer to inner ring. C: Population structure analysis based on ancestral estimation and phylogenetic relationships.

delineate the genetic landscape, we constructed a phylogenetic tree based on SNP data using SNPhylo v.20180901 (Lee et al., 2014), with 1 000 bootstrap replicates. For the identified subpopulations, we calculated the F_{ST} and ratio of genetic polymorphisms (θ_π) using VCFtools v.0.1.17 (Danecek et al., 2011), with parameters set to -fst-windows-size 50 000 and -fst-windows-step 10 000. Regions exhibiting the top 2% of F_{ST} values and the top 5% of the largest and smallest θ_π ratios were flagged as potentially under selection (Wang et al., 2022). Genes within these regions were selected for Gene Ontology (GO) enrichment analysis, with GO terms achieving P<0.05 deemed significantly enriched.

We further explored linkage disequilibrium (LD) decay and long-term effective population size (*Ne*). LD decay was evaluated by plotting the average squared genotypic correlation coefficient (r^2) against 50 kb physical intervals across 24 chromosomes using PopLDdecay v.3.41 (Zhang et al., 2019). The LD decay rate, characterized by the physical distance at which r^2 decreased to half of its maximum value, was determined (Lam et al., 2011). The temporal dynamics of *Ne* were determined using the pairwise sequential Markovian coalescent (PSMC) method, a tool for inferring historical population size changes (Li & Durbin, 2011). The parameters for PSMC included -N25 -t15 -r5 -p "4+25*2+4+6". A mutation rate of 4×10^{-9} per base pair (bp) per generation was assumed, with a generation time (g) of 4 years serving as the basis for analysis.

Principal component analysis (PCA) of individuals in all six groups was performed using PLINK v.1.90b6.21 (Chang et al., 2015). The kinship values of the 226 individuals spanning all six groups were calculated using genome-wide complex trait analysis (GCTA) software v.1.94.0 (Yang et al., 2011).

SNP calling and construction of haplotype reference panel

Raw FASTQ data were filtered using Trimmomatic v.3.29 (Bolger et al., 2014) to eliminate reads of insufficient quality based on the following criteria: (a) reads with quality parameters (Q) below 30, (b) sequences containing partial linker dimer sequences, and (c) reads less than 36 bp. The resulting filtered and clean reads were subsequently aligned to the *P. leopardus* genome assembly (Wang et al., 2020) using

BWA v.0.7.17-r1188 with the Burrows-Wheeler Transform (MEM) algorithm (Li & Durbin, 2010). Following alignment, the data format was converted using SAMtools v.1.12 (Li et al., 2009), and genotype calling was performed using GATK v.4.1.9.0 (McKenna et al., 2010). Further data refinement was performed through filtering using VCFtools v.0.1.17 (Danecek et al., 2011) and GATK, applying the following stringent criteria: Qual By Depth (QD)<2.0, RMS Mapping Quality (MQ)<40, Fisher Strand (FS)<60, Strand Odds Ratio (SOR)>3.0, Mapping Quality Rank Sum Test Allele (MQRankSum)<−12.5, Read Pos Rank Sum Test (ReadRankSum)<−8.0, minor allele frequency (MAF)<0.05, and missing rate>0.1. Finally, SHAPEIT v.4 was employed to pre-phase the called genotypes for each chromosome, enhancing the imputation accuracy when using this haplotype reference panel (Delaneau et al., 2012).

Evaluation of imputation accuracy

Using the haplotype reference panel, the imputation process was executed using GLIMPSE (Rubinacci et al., 2021). To evaluate the accuracy of imputation, a subset of high-coverage resequencing data was randomly selected from the six distinct groups to serve as a validation dataset. Subsequently, low-coverage resequencing data were simulated by systematically downsampling the high-coverage reads to 13 varying coverage depths (0.05×, 0.1×, 0.2×, 0.3×, 0.4×, 0.5×, 0.6×, 0.8×, 1×, 2×, 3×, 4×, and 8×). At each depth, genotype likelihoods were computed. The *P. leopardus* haplotype reference panel was used during the imputation process. Finally, imputation accuracy was assessed by quantifying the squared Pearson correlation (r^2) between the validation dataset and imputed genotypes, following Rubinacci et al. (2021).

GWAS analysis of growth traits

Approximately 8 000 juvenile fish originating from DF were collected and raised, starting from 10 days post-hatching. These individuals were uniformly reared under controlled conditions, with population densities maintained at 50–60 individuals/m^3, ensuring an ample supply of food. After 180 days of growth, 80 large groupers (LG) and 80 small groupers (SG) were carefully selected for sequencing. During the selection process, various morphometric measurements, including body weight (BW), body length (BL), body height (BH), and body thickness (BT), were recorded. Of the 160 individuals, 60 were subjected to high-coverage sequencing (minimum depth of ≥20×), while 100 were subjected to low-coverage sequencing (minimum depth of ≥5×). One individual was excluded from analysis due to sequencing failure. The recorded growth traits (BW, BL, BH, and BT) revealed significant disparities between the LG and SG groups (Supplementary Table S1). Additionally, strong correlations were observed among these growth traits (Supplementary Figure S1). Thus, these growth traits were collectively treated as binary traits for subsequent GWAS analysis, with LG assigned a value of 1 and SG assigned a value of 0.

To enhance the statistical power of GWAS, genotype imputation was employed for low-coverage WGRS data by referencing the haplotype reference panel. To maintain imputation accuracy, individuals with WGRS data volumes below 4.26 Gb (equivalent to a depth of approximately 5×) were systematically excluded from analysis. Ultimately, a dataset of 150 individuals (80 LG and 70 SG) was employed for GWAS analysis, employing linear mixed models in conjunction with genome-wide efficient mixed-model association (GEMMA) software (Zhou & Stephens, 2012). To account for population structure and potential confounding factors, we included a kinship matrix and covariates with the first five PCA values. Ancestral estimations were evaluated to confirm the suitability of the population for GWAS. The number of genetic clusters (*K*-value) was subjected to rigorous testing, ranging from 1 to 10, using ADMIXTURE v.1.3.0 (Alexander et al., 2009). Given the limitations imposed by SNP imputation and LD, we set the GWAS threshold to $P<1\times10^{-5}$, which is more appropriate than the overly stringent Bonferroni test threshold (0.05/*N*), where *N* denotes the number of SNP loci (He et al., 2021). To evaluate the phenotypic variance explained (PVE) by a significant SNP, the following calculation was adopted, as described by Shim et al. (2015):

$$\frac{2\hat{\beta}^2 \mathrm{MAF}(1-\mathrm{MAF})}{2\hat{\beta}^2 \mathrm{MAF}(1-\mathrm{MAF})+(\mathrm{se}(\hat{\beta}))^2 2N\mathrm{MAF}(1-\mathrm{MAF})} \quad (1)$$

where $\hat{\beta}$ is the effect value of the GWAS, *MAF* is the minor allele frequency, *se*($\hat{\beta}$) is the standard error of $\hat{\beta}$, and *N* is the number of individuals.

Candidate gene analysis and real-time quantitative PCR (qPCR) analysis

Candidate genes within a genomic range of 50 kb upstream and 50 kb downstream of the target loci were identified (Wang et al., 2021). Subsequently, the expression profiles of these candidate genes in both the LG and SG populations of *P. leopardus* were analyzed based on qPCR. Primers for the candidate genes and internal reference gene *b2m* (Chen et al., 2021) were designed and synthesized using Integrated DNA Technologies (http://sg.idtdna.com/pages/home) (Supplementary Table S2). During the experimental phase, white muscle tissues were extracted from nine individuals (6 months old) each from the LG and SG groups, derived from an independent population. Equal amounts of muscle tissue were dissected from three individuals within the same group to form a composite sample. Total RNA was extracted using TRIzol reagent (Invitrogen, USA), and cDNA templates were synthesized from 2 μg of RNA using a PrimeScript™ RT Reagent Kit (TaKaRa, Japan). The qPCR analyses were performed using SYBR Green qPCR mix (TaKaRa, Japan) and conducted using a LightCycler 480 (Roche, Switzerland). The thermal cycling conditions comprised an initial denaturation step at 95°C for 5 min, followed by 45 cycles of denaturation at 95°C for 15 s, and annealing/extension at 60°C for 45 s. The relative expression levels of candidate genes were calculated utilizing the $2^{-\Delta\Delta Ct}$ comparative Ct method (Livak & Schmittgen, 2001). Statistical analysis of differences in gene expression levels between the LG and SG populations was conducted using the *t*-test, with significance at $P<0.05$. To ensure robustness and reliability, each experimental procedure was performed in triplicate.

Allele frequency analysis of growth-related SNPs

According to GWAS outcomes, the allele frequencies of growth-related SNPs were calculated using VCFtools v.0.1.17 (Danecek et al., 2011). The criteria for designating an allele as "growth-favorable" were established based on the comparison of allele frequencies between the LG and SG populations. Specifically, if the allele frequency of a growth-related SNP was higher in the LG group compared to the SG group, that allele was deemed as "growth-favorable". To further refine the

selection of candidate SNPs suitable for genetic selection breeding within various clusters, we compared allele frequencies of growth-favorable alleles across eight distinct clusters relative to the LG population. Specifically, SNPs exhibiting an allele frequency ratio in any cluster compared to LG greater than one were identified as potential candidates for genetic selection breeding. Consequently, clusters with a higher abundance of such candidate SNPs were identified, paving the way for future endeavors in selective breeding.

RESULTS

Characterization of SNPs and population structure

A substantial dataset of 11.5 billion reads was obtained from the sequencing of 326 individuals. Following stringent quality control measures, a dataset encompassing 4.14 Tb of clean data was employed for SNP calling. The average alignment rate of the reads to the reference genome reached 98%. This yielded an average sequencing depth of 27.3× for individuals subjected to high-coverage sequencing and 6.2× for those subjected to low-coverage sequencing. A total of 10 926 503 SNPs were initially identified and subsequently refined into a set of 8 735 699 high-quality SNPs for further analysis. Following annotation, most SNPs were located within the intergenic (43.95%) and intronic regions (30.17%), with 12.17% located in upstream regions and 11.47% located in downstream regions. SNPs within coding regions resulted in 32.40% missense mutations, 0.22% nonsense mutations, and 67.379% silent mutations (Table 1).

The distribution of SNPs across the 24 chromosomes was non-uniform. The highest number of SNP loci (458 347) was observed in Chr. 6, while the lowest number (230 591) was found in Chr. 24. Across the whole genome, average SNP density on the chromosomes was approximately 10.00 SNPs/kb, with variations observed among chromosomes. Notably, Chr. 24 showed the highest density (12.40 SNPs/kb) and Chr. 5 showed the lowest density (7.44 SNPs/kb) (Figure 1B).

Based on ancestral estimation analysis with a *K*-value of 8 (Supplementary Figure S2) and phylogenetic relationships, the *P. leopardus* population was effectively divided into eight primary clusters, each characterized by distinct population compositions (Figure 1C). Furthermore, kinship structure analyses and PCA revealed weak genetic relatedness among individuals in the six groups (kinship value<0.5), with no discernible clustering (Supplementary Figure S3).

Table 1 SNP annotation by genomic region, impact, and functional class

Category	Type	Count	Percentage (%)
Genomic region	Downstream	1 197 252	11.468
	Exon	209 114	2.003
	Intergenic	3 149 991	43.951
	Intron	3 149 991	30.173
	Splice site acceptor	91	0.001
	Splice site donor	142	0.001
	Splice site region	24 117	0.231
	Upstream	1 270 749	12.172
Impact	High	902	0.009
	Low	163 957	1.570
	Moderate	68 605	0.657
	Modifier	10 206 390	97.764
Functional class	Missense	68 831	32.404
	Nonsense	459	0.216
	Silent	143 122	67.379

Population genetic characterization and demographic history

Analysis of LD decay revealed notable discrepancies among the eight clusters. Specifically, clusters 1, 3, and 4 exhibited lower LD decay, while cluster 2 displayed the highest LD among the clusters (Figure 2A). In cluster 2, the LD level (r^2) decreased to half of its maximum value (0.3) at a physical distance of 72 bp, whereas this decline occurred at 26 bp in cluster 4. The temporal dynamics of *N*e over the past 10^6 years were elucidated (Figure 2B). Analysis indicated that the *N*e of *P. leopardus* reached its lowest point approximately 1×10^5 years ago. From two million years ago (Ma) to 100 kilo-years ago (ka), the *N*e of *P. leopardus* steadily declined, followed by a divergence in *N*e values of the eight clusters starting at 100 ka.

Analysis of the genotypes of SNP loci in specific regions revealed a significant divergence among clusters (F_{ST}>0.25). Notably, distinct regions were identified in clusters 2, 5, 6, 7, and 8. In cluster 2, these regions were located on Chr. 15 and Chr. 16. These regions were distributed on two chromosomes in cluster 5 (Chr. 7 and Chr. 9), but on five chromosomes in cluster 6 (Chr. 9, Chr. 11, Chr. 12, Chr. 16, and Chr. 21) and three chromosomes in cluster 7 (Chr. 12, Chr. 15, and Chr. 17). Cluster 8 harbored the highest number of specific regions, dispersed across Chr. 4, Chr. 6, Chr. 13, Chr. 14, Chr. 18, and Chr. 19 (Figure 2C).

The pairwise F_{ST} values between each pair of clusters ranged from 0.018785 to 0.116810. While genetic stratification was observed in clusters 5, 6, 7, and 8 (F_{ST}>0.05), clusters 1, 3, and 4 demonstrated a relatively high degree of genetic similarity (Figure 2D).

Genome selection signatures in clusters

Selective sweep analysis based on F_{ST} values and θ_π score ratios revealed selected genomic regions in clusters 5, 6, 7, and 8. In these clusters, 246, 431, 361, and 578 candidate windows of strong selection were identified, containing 164, 280, 229, and 342 genes, respectively (Figure 3A; Supplementary Figure S4 and Table S3). GO enrichment analysis elucidated the functional roles of these selective genes. In cluster 5, selective genes were primarily enriched in processes such as diphthine synthase activity, brown fat cell differentiation, and 5-3 RNA helicase activity. In cluster 6, selective genes were enriched in thiamine metabolic processes, ammonium transmembrane transporter activity, and semaphorin receptor binding, while those in cluster 7 were enriched in GABA-A receptor activity, S-(hydroxymethyl) glutathione dehydrogenase activity, and Schwann cell differentiation, and those in cluster 8 were enriched in olfactory receptor activity and xenobiotic transporter activity (Figure 3B).

Imputation performance based on haplotype reference panel

The *P. leopardus* haplotype reference panel was constructed as shown in Figure 4A, which contained information from 8 735 699 high-quality SNP loci. Notably, with increasing WGRS data coverage depth, there was a continuous improvement in imputation accuracy (r^2), supported by the

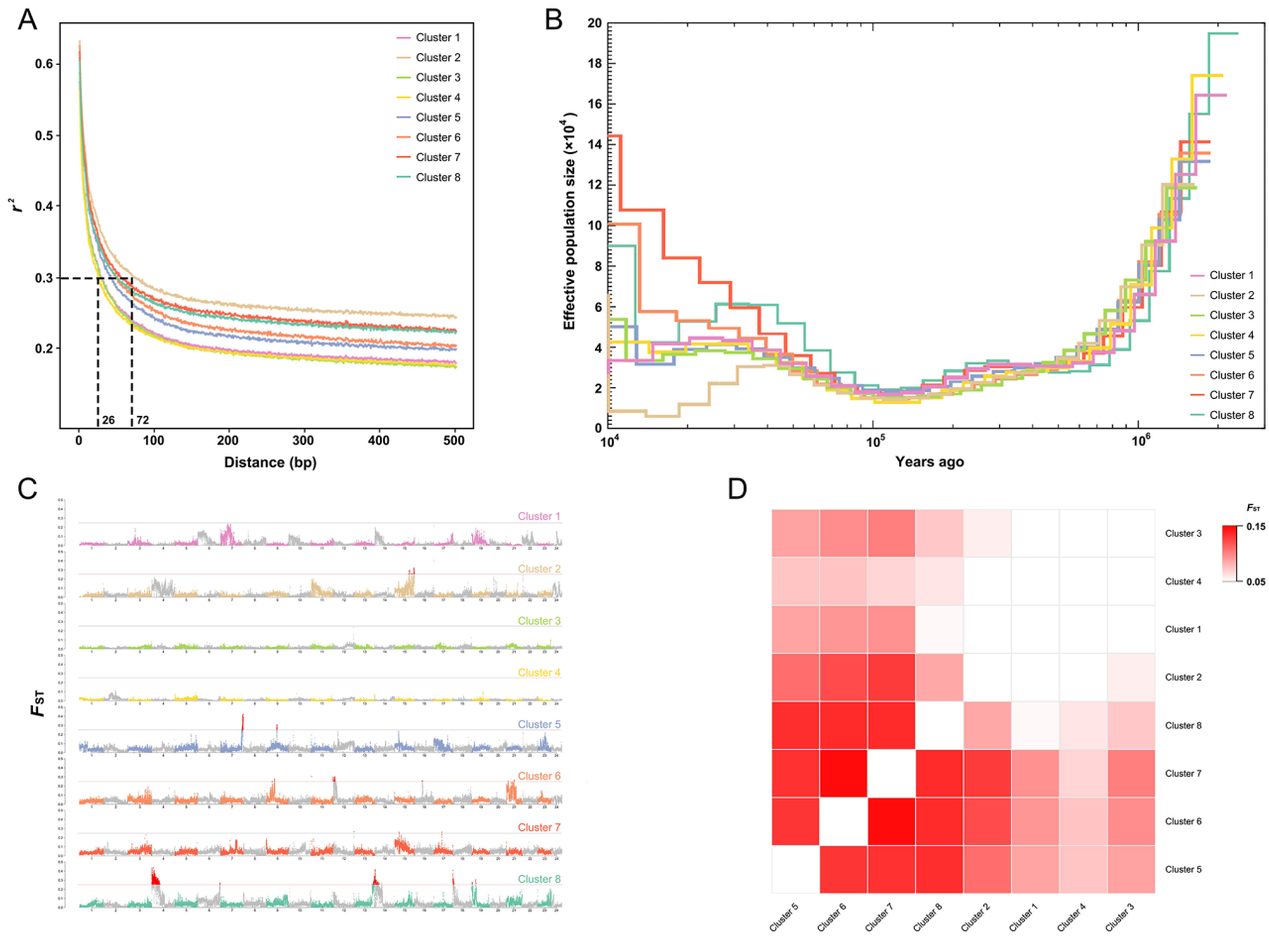

Figure 2 Population genetic analysis of eight clusters

A: LD decay at 0–500 bp. B: Changes in *Ne* over the past ~10^6 years in eight clusters. C: F_{ST} values between one cluster and other clusters were plotted against the positions on each of the 24 chromosomes. Plots marked in red represent divergence region between two populations (F_{ST}>0.25). D: Pairwise F_{ST} matrix of eight clusters.

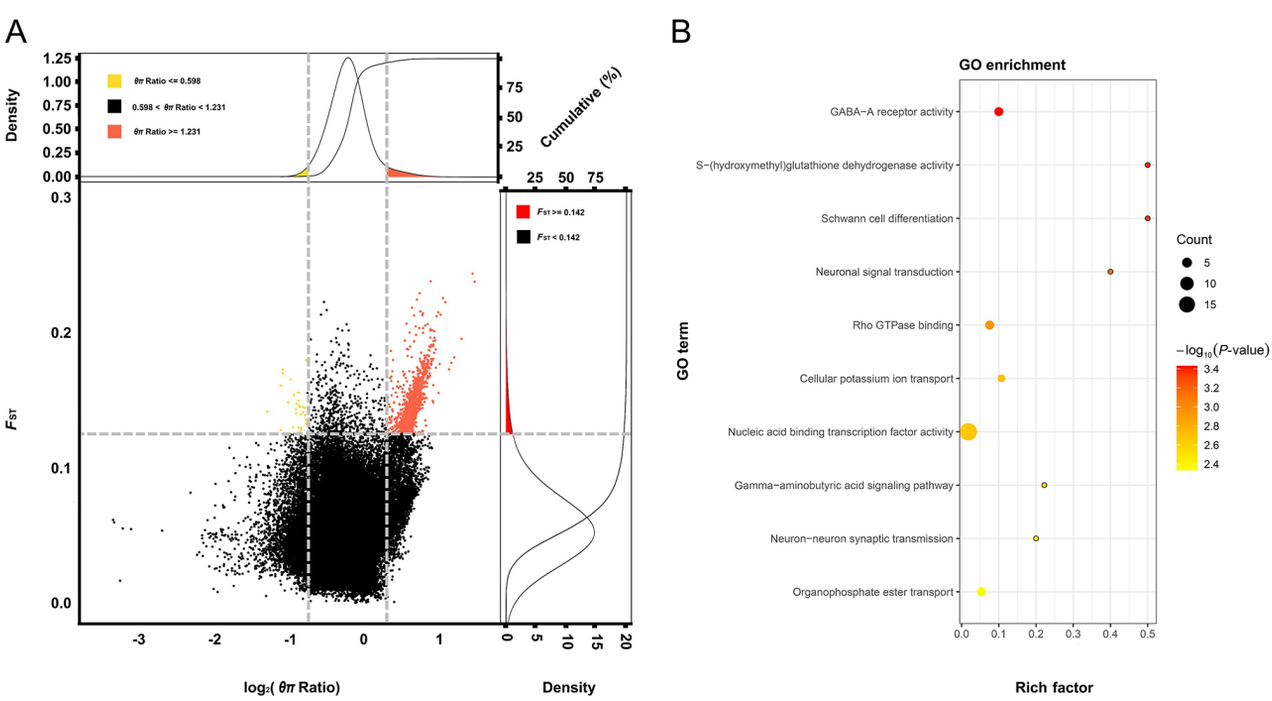

Figure 3 Selective sweep and GO enrichment analysis of cluster 7

A: Distribution of *θπ* ratios (*θπ* reference population/*θπ* selective cluster) and F_{ST} values calculated in 50 kb window with a 10 kb step. Windows with 5% largest or smallest *θπ* ratios and top 2% largest F_{ST} values were identified as regions under selection. B: Enriched GO terms of genes under selection.

haplotype reference panel. This trend was consistent across the entire genome. Remarkably, even at a minimal coverage depth of 0.05×, imputation accuracy reached 0.82, increasing to 0.95 at relatively low coverage depths of 4× (Figure 4B).

Growth-related SNPs and candidate genes

Constructing a high-quality haplotype reference panel enabled the use of low-coverage GWAS data. A total of 7 651 790 SNP loci were obtained for imputation analysis. Comparatively, the imputed data retained an additional 1 724 806 SNP loci compared with the original dataset.

In the GWAS of growth traits, Manhattan and quantile-quantile plots were used to visualize the results (Figure 5A, B). Notably, 75 growth-related SNP loci spanning 18 chromosomes reached genome-wide significance ($P<1\times10^{-5}$) (Supplementary Table S4). The most significant SNP (Chr17:22316930) was located on Chr. 17 (P=9.44e-8). Proximate to this locus, an 800 kb region was enriched with growth-related SNP loci. This region contained seven candidate genes, including myocyte enhancer factor 2D (*mef2d*), transmembrane protein 79a (*tmem79a*), S100 calcium-binding protein S (*s100s*), pre-B-cell leukemia transcription factor-interacting protein (*pbxip1*), cyclin-dependent kinase regulatory subunit 1b (*cks1b*), ephrin-A3 (*efna3*), and tripartite motif-containing 46b (*trim46b*) (Figure 5C). The SNP (Chr9:3779548) with the highest PVE (5.28%) was located on chromosome 9. Genotype composition analysis of the three most significant growth-related SNPs based on *P*-values and PVE revealed substantial differences in genotypes between LG and SG (Supplementary Figure S5). Kinship matrix analysis indicated that the genetic relatedness among most individuals was weak (kinship value<0.5) and the cross-validation error reached its lowest point at K=6 (Supplementary Figure S6). These results suggest that the experimental population used for GWAS could be considered as a natural population with a rich genetic foundation.

In total, 49 candidate genes were identified within a 50 kb region upstream and downstream of the growth-related SNPs (Supplementary Table S5). These genes are primarily involved in growth-related pathways, including disintegrin and metalloproteinase domain-containing protein 12 (*adam12*), SH3 and PX domain-containing protein 2A (*sh3pxd2a*), growth arrest-specific protein (*gas7*), semaphorin-3ab (*sema3ab*), semaphorin-3D (*sema3d*), and palmitoyltransferase ZDHHC17 (*zdhhc17*). Relative expression analysis of the candidate genes in LG and SG demonstrated that three genes, *mef2d*, sorting nexin-22 (*snx22*), and *cks1b*, were significantly up-regulated in the LG group (Supplementary Figure S7).

Allele frequencies of growth-related SNPs in clusters

Within the set of 75 growth-related SNPs identified through GWAS analysis, 68 were located within eight clusters. Subsequently, allele frequencies for these SNP loci were calculated for each cluster (Supplementary Table S6). Given that some SNPs exhibited high LD levels, with r^2 values greater than 0.1, 43 growth-related SNPs were retained for further analysis. These SNPs exhibited different distributions among the clusters. Clusters 1, 2, 3, and 4 contained five, seven, six, and two candidate SNPs, respectively, primarily on chromosome 22. Clusters 6, 7, and 8 exhibited a more scattered distribution, with nine, nine, and three candidate SNPs, respectively. Cluster 5 contained the highest number of candidate SNPs, with a high proportion of growth-favorable alleles (12). Of note, one of the top three most significant SNPs (Chr17:21537423) based on *P*-value and one of the top three most significant SNPs (Chr10:29219664) based on PVE were detected in cluster 5.

DISCUSSION

Population genetic analysis based on cultured groups has important practical implications for germplasm resource management and breeding efforts (Zhang et al., 2018a). Our investigation revealed a complex situation in cultured *P. leopardus* groups, characterized by a lack of specific data on germplasm sources and the intermixing of parental populations from different origins to generate offspring. To optimize the use of *P. leopardus* germplasm resources, we performed population structure analysis using ancestral estimation and phylogenetic analysis (Figure 1C), both reliable methods for population structure studies (Khadka et al., 2020). Given the lack of systematic selective breeding and the reliance on crossbreeding of parental populations for producing offspring, leading to no distinct *P. leopardus* families, we speculate that the genetic differences among the eight clusters may be attributed to mutations inherited from ancestral individuals. These mutations in wild ancestral individuals likely underwent adaptive evolution, a central thesis in evolutionary biology suggesting that naturally selected heritable phenotypic traits lead to adaptation and differentiation among populations inhabiting environments with distinct selective regimes (Linhart & Grant, 1996). Our population genetic analysis of the eight clusters aimed to shed light on the genetic characteristics inherited from ancestors.

LD decay analysis has been widely used to understand evolutionary and demographic events (Esteras et al., 2013). The present study revealed variations in LD among the eight clusters, with clusters 2, 5, 6, 7, and 8 exhibiting higher LD levels than clusters 1, 3, and 4 (Figure 2A). Differences in LD can be attributed to various factors, such as admixture, mutations, and changes in *Ne* (Slatkin, 2008). The variations in LD among the eight clusters may also result from adaptive evolution in ancestral individuals, as observed in other species. For example, fitness-affecting mutations in honeybees (*Apis mellifera*) have been shown to alter LD patterns in the genome (Kent et al., 2011). Similarly, extensive LD and parallel adaptive divergence have been observed in the genomes of threespine sticklebacks (*Gasterosteus aculeatus*) (Hohenlohe et al., 2012). Fluctuations in *Ne* further support this hypothesis. Notably, divergence in *Ne* among the eight clusters began around 100 ka (Figure 2B), coinciding with the end of the Middle Pleistocene, a period marked by a global drop in sea level (Auer et al., 2021), potentially promoting population diffusion and differentiation. The *Ne* of *P. leopardus* declined sharply from 2 Ma until it reached its minimum at approximately 100 ka (Figure 2B). This decline has been observed in other marine organisms, such as coral (*Acropora millepora*) (Fuller et al., 2020) and limpets (*Bathyacmaea lacteal*) (Liu et al., 2020), and may be related to marine life extinction events around the Pliocene-Pleistocene boundary (Tomiyama et al., 2020), indicating that the ecosystems of *P. leopardus* were also influenced by the paleoclimate.

Identification of markers for cluster differentiation can enhance the effective management of germplasm resources. Specific regions characterized by substantial genetic divergence were identified within the clusters through F_{ST}

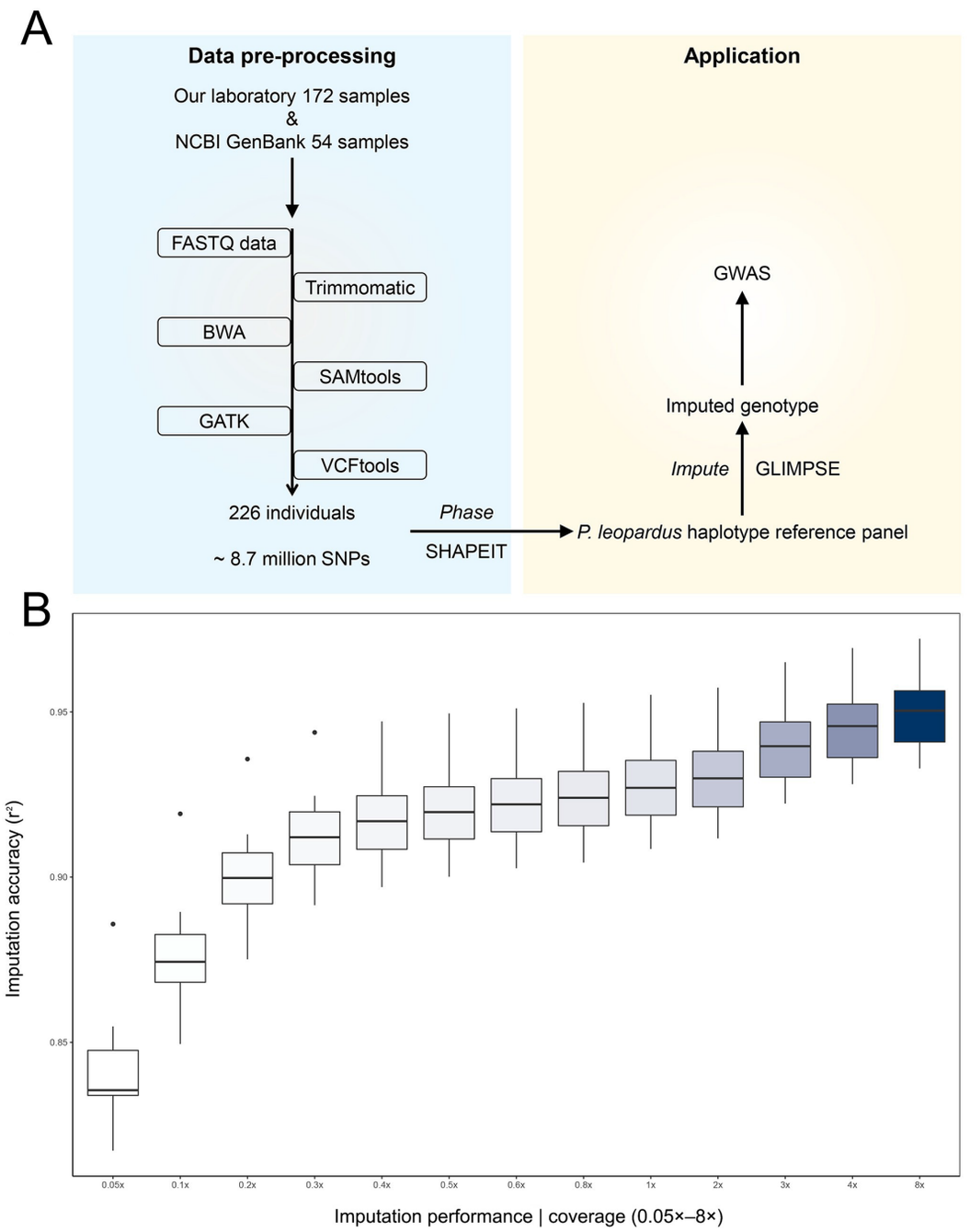

Figure 4 Construction and imputation performance of haplotype reference panel

A: Flow chart of haplotype reference panel construction and application. B: Imputation accuracy for resequencing data at different depths.

analysis (Figure 2C). The genetic diversity elucidated by F_{ST} is crucial for establishing the validity of DNA fingerprints (Khan et al., 2021). These regions within the clusters offer potential for creating DNA fingerprints to determine the cluster affiliation of individuals. DNA fingerprints based on SNP data have been widely used to distinguish subpopulations in various crops, including cucumber (Guo et al., 2019) and maize (Zhang et al., 2020). SNPs have also emerged as promising tools for genetic characterization in aquaculture, particularly in the context of fingerprinting (Amoussou et al., 2019). Notably, several regions harboring a multitude of stratified SNP loci were identified within the eight clusters, characterized by F_{ST} values exceeding 0.25. These regions may serve as valuable resources for the development of DNA fingerprints, thereby enhancing efficient management of germplasm resources.

In accordance with Wright's guidelines (Wright, 1949), our analysis revealed populations stratifications in four specific clusters (clusters 5, 6, 7, and 8), evidenced by F_{ST} values surpassing 0.05 (Figure 2D). This suggests that certain selective loci within these four clusters may facilitate adaptation to the environmental conditions experienced by *P. leopardus*, thereby offering pathways to genetic improvement. F_{ST} values and θ_π score ratios have been widely used to detect selective sweeps and genetic differentiation (Li et al., 2013; Wang et al., 2022). This approach revealed several potential elite phenotypic traits in *P. leopardus*. For instance, the *ppargc1a* gene in cluster 5 is known to influence islet insulin secretion (Ling et al., 2008) and individual variability in fat loss in humans (Mazur et al., 2020), suggesting a potential role in modulating insulin secretion and fat metabolism in *P.*

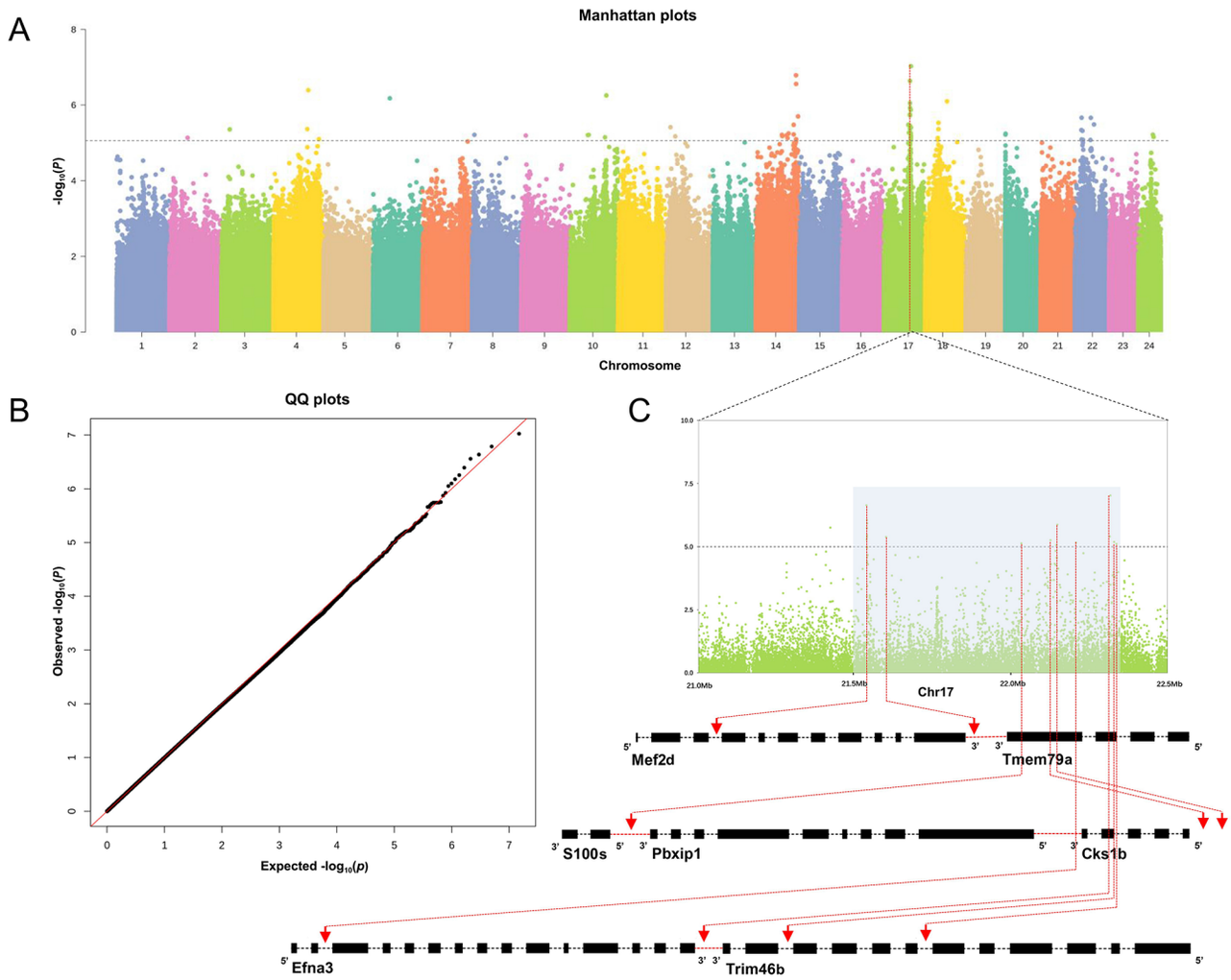

Figure 5 GWAS analysis for growth traits

A: Manhattan plots of GWAS analysis for growth traits. B: QQ plots of GWAS analysis for growth traits. Horizontal dotted line indicates genome-wide significance level (-$\log_{10}$(*P*-value)>5). C: Significant region associated with growth traits. Growth-related SNPs and genes are marked.

leopardus. Similarly, the identification of the *Ralgapa1* gene in clusters 5 and 7, associated with neurodevelopmental disabilities and muscular hypotonia in humans (Wagner et al., 2020), may contribute to improved food acquisition competitiveness. Multiple olfactory receptor genes, such as olfactory receptor 15 (*olfr15*), olfactory receptor 49 (*olfr49*), and olfactory receptor 10J5 (*or10j5*), were selected in cluster 8. The *olfr15* gene is activated by (methylthio) methanethiol (MTMT) in various species, including dogs (*Canis lupus familiaris*), pigs (*Sus scrofa*), rabbits (*Oryctolagus cuniculus*), hamsters (*Mesocricetus auratus*), and rats (*Rattus norvegicus*) (Zhang et al., 2018b), and is also responsive to odorants such as acetophenone and benzaldehyde (Strotmann et al., 2004). In mice (*Mus musculus*), *olfr49* responds to (-)-citronellal (Krautwurst et al., 1998), whereas OR10J5 can be activated by lyral (Grosmaitre et al., 2006). These four genes may significantly enhance appetite and food capture capabilities by augmenting sensitivity to olfactory perception. In summary, the identified selective signatures offer valuable insights into future breeding directions.

Population genetic analysis revealed multiple genotypes representing distinct genetic groups in the WGRS data. We constructed the first haplotype reference panel for *P. leopardus* using 226 high-coverage WGRS datasets, which facilitated high-precision imputation. Previous research on humans has demonstrated that genotypes imputed from low-pass WGRS data yield increased power for GWAS compared with genotypes imputed from array data (Wasik et al., 2021). In *S. scrofa*, imputation of very low-coverage data (1.11×) using the GLIMPSE method demonstrates superior performance (Nosková et al., 2021). In fish species, low-density SNP data imputation has been employed in GWAS to identify SNP loci associated with critical traits (Bargelloni et al., 2021). WGRS imputation, underpinned by a robust haplotype panel and imputation algorithm, represents a cost-effective means of obtaining precise genotypic information. Thus, our study offers a reliable *P. leopardus* haplotype reference panel for future application in diverse breeding programs targeting various traits, such as growth and disease resistance.

The imputed WGRS data served as the foundation for GWAS targeting growth traits, resulting in the detection of 75 growth-related SNPs spanning 18 chromosomes in *P. leopardus*. The identification of multiple growth-related SNPs extends to other grouper species, indicating the polygenetic nature of growth traits within the grouper family. For instance, 23 growth-related SNPs across 12 chromosomes have been identified in the brown-marbled grouper (Yang et al., 2020b), while 36 growth-related SNPs distributed on 14 chromosomes have been identified in the giant grouper (Wu et al., 2019).

Similarly, in the orange-spotted grouper, 88 SNPs associated with body weight, spanning 21 chromosomes, have been reported (Xu et al., 2019). Among the identified growth-related SNPs, a significant region containing seven candidate genes and conserved synteny has emerged, as evidenced by the evolutionary conservation of *cks1b*, *efna3*, *trim46a*, and *s100s* in *H. sapiens*, platy fish (*Xiphophorus maculatus*), and *P. leopardus* (Supplementary Figure S8). Previous studies have underscored the pivotal roles of these genes in the growth process (Liang et al., 2017; Most et al., 2013; Wang et al., 2014; Zeng et al., 2019). Notably, several of these genes and their homologs, such as *Adam12* in catfish (Li et al., 2018), have been implicated in growth traits across various species.

Three candidate genes (*mef2d*, *cks1b* and *snx22*) exhibited differential expression levels between the LG and SG groups. In the African clawed frog (*Xenopus laevis*), *mef2d* plays a role in dermomyotome formation and lateral myogenesis (Gaspera et al., 2013). Additionally, down-regulation of *cks1b* in retinoblastoma cells inhibits cell proliferation and migration through the MEK/ERK signaling pathway (Zeng et al., 2019). These findings suggest that *mef2d* and *cks1b* may regulate cell proliferation and differentiation in skeletal muscle, thereby promoting individual growth. Furthermore, although *snx22* has been proposed as a potential radioresistant gene in nasopharyngeal carcinoma (Zhou et al., 2023), its precise role in growth-related processes requires further investigation. Collectively, these results underscore the robustness and reliability of GWAS focused on growth traits in *P. leopardus*.

Investigation of allele frequencies is well-established as a reliable approach for genetically selective breeding. In aquaculture, SNPs associated with specific traits, identified through GWAS, have been increasingly used for genetic selective breeding in various species, leveraging favorable alleles linked to specific traits (Correa et al., 2017; Yue, 2014). Presently, the implementation of one-on-one fertilization during spawning events in *P. leopardus* remains challenging. A more viable strategy involves the utilization of groups harboring a high proportion of growth-favorable alleles for breeding purposes. Notably, cluster 5 exhibited a significant presence of growth-favorable alleles among the identified candidate SNPs (Figure 6). In future selective breeding programs focused on growth traits of *P. leopardus*, extensive utilization of allele resources from cluster 5 could play a pivotal role in enhancing growth characteristics of this species.

CONCLUSIONS

Through comprehensive material collection efforts in Hainan, a substantial dataset of WGRS data for *P. leopardus* has been accumulated, comprising more than 8.73 million high-quality SNP loci, which have been crucial for population structure analysis and haplotype reference panel construction. The establishment of this reference panel holds significant potential for reducing sequencing costs in future GWAS, thereby facilitating advancements in breeding initiatives. Analysis of these data revealed the presence of eight distinct clusters among the sampled individuals, with clear stratifications within four clusters. Selective sweep analysis identified 1 614 genes within these four clusters, underscoring their potential importance for future breeding efforts. Using the haplotype reference panel, precise low-coverage WGRS imputation was successfully performed for GWAS purposes. Notably, analysis of allele frequencies for growth-related SNPs across the eight clusters revealed that cluster 5 harbored the highest number of candidate SNPs suitable for genetic selective breeding programs. This research marks a significant advancement in genome-wide studies of *P. leopardus* and provides a promising path for molecular breeding initiatives. In conclusion, this study provides a robust tool and valuable insights for managing germplasm resources and fostering genome-associated breeding efforts for *P. leopardus*.

DATA AVAILABILITY

The genome resequencing data were deposited in the NCBI (PRJNA955737), Science Data Bank (10.57760/sciencedb.14280), and GSA (PRJCA022173) databases.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

J.J.H. and Z.M.B., conception and design of the study; S.X.W. and W.T.H., analysis and interpretation of data; H.D., Meng-Ya W., Ming-Yi W., M.X.T., P.Y.L., and X.G., acquisition of data; S.X.W., drafting of the article; B.W. and Q.F.Z., revising the manuscript for important intellectual content. J.J.H. and Z.M.B., final approval of the version to be submitted. All authors read and approved the final version of the manuscript.

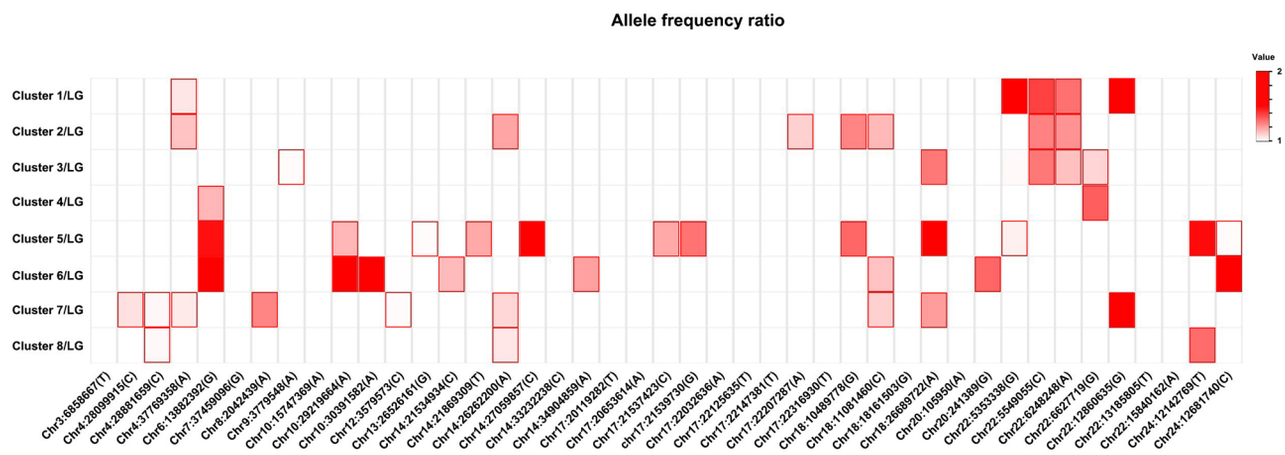

Figure 6 Allele frequency ratios of clusters to LG

Candidate SNPs for genetic selective breeding are in red.

ACKNOWLEDGMENTS

We would like to thank colleagues from Hainan Chenhai Aquatic Co., Ltd. (China) for their help in collecting samples.

REFERENCES

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**(9): 1655−1664.

Amoussou TO, Youssao Abdou Karim I, Dayo GK, et al. 2019. An insight into advances in fisheries biology, genetics and genomics of african tilapia species of interest in aquaculture. *Aquaculture Reports*, **14**: 100188.

Auer G, Petrick B, Yoshimura T, et al. 2021. Intensified organic carbon burial on the australian shelf after the middle pleistocene transition. *Quaternary Science Reviews*, **262**: 106965.

Bargelloni L, Tassiello O, Babbucci M, et al. 2021. Data imputation and machine learning improve association analysis and genomic prediction for resistance to fish photobacteriosis in the gilthead sea bream. *Aquaculture Reports*, **20**: 100661.

Bielikova OY, Mariutsa AE, Mruk AI, et al. 2021. Genetic structure of rainbow trout oncorhynchus mykiss (Salmoniformes, Salmonidae) from aquaculture by DNA-markers. *Biosystems Diversity*, **29**(1): 28−32.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, **30**(15): 2114−2120.

Chang CC, Chow CC, Tellier LCAM, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**(1): 7.

Chen XJ, Sun Y, Zhang PP, et al. 2021. Screening of stable internal reference genes by quantitative real-time PCR in humpback grouper *Cromileptes altivelis*. *Journal of Oceanology and Limnology*, **39**(5): 1985–1999.

Correa K, Lhorente JP, Bassini L, et al. 2017. Genome wide association study for resistance to *Caligus rogercresseyi* in atlantic salmon (*Salmo Salar* L.) using a 50K SNP genotyping array. *Aquaculture*, **472**(S1): 61−65.

Danecek P, Auton A, Abecasis G, et al. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15): 2156−2158.

Delaneau O, Marchini J, Zagury JF. 2012. A linear complexity phasing method for thousands of genomes. *Nature Methods*, **9**(2): 179−181.

Esteras C, Formisano G, Roig C, et al. 2013. SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium. *Theoretical and Applied Genetics*, **126**(5): 1285−1303.

Froese R, Pauly D. 2023[2023-02]. FishBase. World Wide Web Electronic Publication.https://www.fishbase.se/search.php/.

Fuller ZL, Mocellin VJL, Morris LA, et al. 2020. Population genetics of the coral *Acropora millepora*: toward genomic prediction of bleaching. *Science*, **369**(6501): eaba4674.

Gaspera BD, Armand AS, Lecolle S, et al. 2013. *Mef2d* acts upstream of muscle identity genes and couples lateral myogenesis to dermomyotome formation in *Xenopus Laevis*. *PLoS One*, **7**(12): e52359.

Grosmaitre X, Vassalli A, Mombaerts P, et al. 2006. Odorant responses of olfactory sensory neurons expressing the odorant receptor MOR23: A patch clamp analysis in gene-targeted mice. *Proceedings of the National Academy of Sciences*, **103**(6): 1970−1975.

Guo ZF, Wang HW, Tao JJ, et al. 2019. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. *Molecular Breeding*, **39**(3): 37.

Han WT, Wu SX, Ding H, et al. 2023. Improved chromosomal-level genome assembly and re-annotation of leopard coral grouper. *Scientific Data*, **10**(1): 156.

He X, Li CY, Qi HG, et al. 2021. A genome-wide association study to identify the genes associated with shell growth and shape-related traits in crassostrea gigas. *Aquaculture*, **543**: 736926.

Hohenlohe PA, Bassham S, Currey M, et al. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **367**(1587): 395−408.

Kent CF, Issa A, Bunting AC, et al. 2011. Adaptive evolution of a key gene affecting queen and worker traits in the honey bee. *Apis Mellifera*. *Molecular Ecology*, **20**(24): 5226−5235.

Kessuwan K, Kubota S, Liu Q, et al. 2016. Detection of growth-related quantitative trait loci and high-resolution genetic linkage maps using simple sequence repeat markers in the kelp grouper (*Epinephelus Bruneus*). *Marine Biotechnology*, **18**(1): 57−84.

Khadka K, Torkamaneh D, Kaviani M, et al. 2020. Population structure of nepali spring wheat (*Triticum Aestivum* L.) germplasm. *BMC Plant Biology*, **20**(1): 530.

Khan MMH, Rafii MY, Ramlee SI, et al. 2021. DNA fingerprinting, fixation-index (Fst), and admixture mapping of selected bambara groundnut (*Vigna Subterranea* [L.] Verdc.) accessions using ISSR markers system. *Scientific Reports*, **11**(1): 14527.

Krautwurst D, Yau KW, Reed RR. 1998. Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell*, **95**(7): 917−926.

Lam HM, Xu X, Liu X, et al. 2011. Addendum: Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics*, **43**(4): 387.

Lee TH, Guo H, Wang XY, et al. 2014. SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, **15**(1): 162.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with burrows–wheeler transform. *Bioinformatics*, **26**(5): 589−595.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, **475**(7357): 493−496.

Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078−2079.

Li MZ, Tian SL, Jin L, et al. 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and tibetan wild boars. *Nature Genetics*, **45**(12): 1431−1438.

Li N, Zhou T, Geng X, et al. 2018. Identification of novel genes significantly affecting growth in catfish through GWAS analysis. *Molecular Genetics and Genomics*, **293**(3): 587−599.

Liang Q, Fang JY, Hong J. 2017. Ubiquitin ligase TRIM46 promotes the proliferation and invasion of colorectal cancer cells. *Tumor*, **37**(4): 334−339.

Ling C, Del Guerra S, Lupi R, et al. 2008. Epigenetic regulation of *PPARGC1A* in human type 2 diabetic islets and effect on insulin secretion. *Diabetologia*, **51**(4): 615−622.

Linhart YB, Grant MC. 1996. Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, **27**: 237−277.

Liu HT, Dai C, Li XX, et al. 2022. Analysis of the differentially expressed genes and characterization of *PIMITF* genes of the leopard coral grouper (*Plectropomus Leopardus*). *Aquaculture Reports*, **23**: 101076.

Liu RY, Wang K, Liu J, et al. 2020. De novo genome assembly of limpet *Bathyacmaea Lactea* (Gastropoda: Pectinodontidae): the first reference genome of a deep-sea gastropod endemic to cold seeps. *Genome Biology and Evolution*, **12**(6): 905−910.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods*, **25**(4): 402−408.

Marchini J, Howie B. 2010. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, **11**(7): 499−511.

Mazur II, Drozdovska S, Andrieieva O, et al. 2020. PPARGC1A gene polymorphism is associated with exercise-induced fat loss. *Molecular Biology Reports*, **47**(10): 7451−7457.

McKenna A, Hanna M, Banks E, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**(9): 1297−1303.

Meirmans PG, Hedrick PW. 2011. Assessing population structure: F_{ST} and

related measures. *Molecular Ecology Resources*, **11**(1): 5−18.
Mekuchi M, Asakura T, Sakata K, et al. 2018. Intestinal microbiota composition is altered according to nutritional biorhythms in the leopard coral grouper (*Plectropomus leopardus*). *PLoS One*, **13**(6): e0197256.
Most P, Lerchenmüller C, Rengo G, et al. 2013. S100A1 deficiency impairs postischemic angiogenesis via compromised proangiogenic endothelial cell function and nitric oxide synthase regulation. *Circulation Research*, **112**(1): 66−78.
Nosková A, Bhati M, Kadri NK, et al. 2021. Characterization of a haplotype-reference panel for genotyping by low-pass sequencing in swiss large white pigs. *BMC Genomics*, **22**(1): 290.
Parvez MS, Sarower MG, Hasan MM, et al. 2021. Monitoring of genetic variation in mono-sex tilapia stocks using randomly amplified polymorphic DNA markers in some private hatcheries of bangladesh. *Egyptian Journal of Aquatic Biology and Fisheries*, **25**(1): 1017−1030.
Peter BM. 2016. Admixture, population structure, and *F*-statistics. *Genetics*, **202**(4): 1485−1501.
Rimmer MA, Glamuzina B. 2019. A review of grouper (Family Serranidae: Subfamily Epinephelinae) aquaculture from a sustainability science perspective. *Reviews in Aquaculture*, **11**(1): 58−87.
Rubinacci S, Ribeiro DM, Hofmeister RJ, et al. 2021. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics*, **53**(1): 120−126.
Shim H, Chasman DI, Smith JD, et al. 2015. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 caucasians. *PLoS One*, **10**(4): e0120758.
Slatkin M. 2008. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, **9**(6): 477−485.
Strotmann J, Levai O, Fleischer J, et al. 2004. Olfactory receptor proteins in axonal processes of chemosensory neurons. *The Journal of Neuroscience* , **24**(35): 7754−7761.
Tam V, Patel N, Turcotte M, et al. 2019. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, **20**(8): 467−484.
Tomiyama JM, Takagi D, Kantar MB. 2020. The effect of acute and chronic food shortage on human population equilibrium in a subsistence setting. *Agriculture & Food Security*, **9**(1): 6.
Wagner M, Skorobogatko Y, Pode-Shakked B, et al. 2020. Bi-allelic variants in *RALGAPA1* cause profound neurodevelopmental disability, muscular hypotonia, infantile spasms, and feeding abnormalities. *The American Journal of Human Genetics*, **106**(2): 246−255.
Wang H, Teng MX, Liu PP, et al. 2022. Selection signatures of pacific white shrimp *Litopenaeus vannamei* revealed by whole-genome resequencing analysis. *Frontiers in Marine Science*, **9**: 844597.
Wang SY, Wang HZ, Zhao L, et al. 2021. Identification of genes associated with carotenoids accumulation in scallop (*Patinopecten Yessoensis*). *Aquaculture*, **550**: 737850.
Wang XF, Yang ZJ, Jiang LM, et al. 2023. Assessment of germplasm resource and detection of genomic signature under artificial selection of Zhikong scallop (*Chlamys farreri*). *Aquaculture*, **574**: 739730.
Wang YB, Wen X, Zhang XH, et al. 2020. Chromosome genome assembly of the leopard coral grouper (*Plectropomus Leopardus*) with nanopore and Hi-C sequencing data. *Frontiers in Genetics*, **11**: 876.
Wang ZG, Yin BL, Wang B, et al. 2014. MicroRNA-210 promotes proliferation and invasion of peripheral nerve sheath tumor cells targeting EFNA3. *Oncology Research*, **21**(3): 145−154.
Wasik K, Berisa T, Pickrell JK, et al. 2021. Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genomics*, **22**(1): 197.
Wright S. 1949. The genetical structure of populations. *Annals of Eugenics*, **15**(1): 323−354.
Wu LN, Yang Y, Li BJ, et al. 2019. First genome-wide association analysis for growth traits in the largest coral reef-dwelling bony fishes, the giant grouper (*Epinephelus Lanceolatus*). *Marine Biotechnology*, **21**(5): 707−717.
Xia SD, Sun JH, Li M, et al. 2020. Influence of dietary protein level on growth performance, digestibility and activity of immunity-related enzymes of leopard coral grouper, *Plectropomus Leopardus* (Lacépède, 1802). *Aquaculture Nutrition*, **26**(2): 242−247.
Xu TF, Zhang XH, Ruan ZQ, et al. 2019. Genome resequencing of the orange-spotted grouper (*Epinephelus Coioides*) for a genome-wide association study on ammonia tolerance. *Aquaculture*, **512**: 734332.
Yang J, Lee SH, Goddard ME, et al. 2011. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, **88**(1): 76−82.
Yang Y, Wu LN, Chen JF, et al. 2020a. Whole-genome sequencing of leopard coral grouper (*Plectropomus Leopardus*) and exploration of regulation mechanism of skin color and adaptive evolution. *Zoological Research*, **41**(3): 328−340.
Yang Y, Wu LN, Wu X, et al. 2020b. Identification of candidate growth-related SNPs and genes using GWAS in brown-marbled grouper (*Epinephelus Fuscoguttatus*). *Marine Biotechnology*, **22**(2): 153−166.
Yu H, You XX, Li J, et al. 2018. A genome-wide association study on growth traits in orange-spotted grouper (*Epinephelus coioides*) with RAD-seq genotyping. *Science China Life Sciences*, **61**(8): 934−946.
Yue GH. 2014. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries*, **15**(3): 376−396.
Zeng QF, Zhao BJ, Wang H, et al. 2022. Aquaculture Molecular Breeding Platform (AMBP): a comprehensive web server for genotype imputation and genetic analysis in aquaculture. *Nucleic Acids Research*, **50**(W1): W66−W74.
Zeng Z, Gao ZL, Zhang ZP, et al. 2019. Downregulation of CKS1B restrains the proliferation, migration, invasion and angiogenesis of retinoblastoma cells through the MEK/ERK signaling pathway. *International Journal of Molecular Medicine*, **44**(1): 103−114.
Zhang C, Dong SS, Xu JY, et al. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**(10): 1786−1788.
Zhang HY, Zhao ZX, Xu J, et al. 2018a. Population genetic analysis of aquaculture salmonid populations in China using a 57K rainbow trout SNP array. *PLoS One*, **13**(8): e0202582.
Zhang J, Yang JJ, Zhang LK, et al. 2020. A new SNP genotyping technology target SNP-seq and its application in genetic analysis of cucumber varieties. *Scientific Reports*, **10**(1): 5623.
Zhang JB, Liu HF, Song Y. 2010. Development and characterization of polymorphic microsatellite loci for a threatened reef fish *Plectropomus leopardus*. *Conservation Genetics Resources*, **2**(1): 101–103.
Zhang RN, Pan Y, Ahmed L, et al. 2018b. A multispecific investigation of the metal effect in mammalian odorant receptors for sulfur-containing compounds. *Chemical Senses*, **43**(5): 357−366.
Zhou Q, Guo XY, Huang Y, et al. 2020. De novo sequencing and chromosomal-scale genome assembly of leopard coral grouper. *Plectropomus Leopardus*. *Molecular Ecology Resources*, **20**(5): 1403−1413.
Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**(7): 821−824.
Zhou ZY, Chen G, Shen MJ, et al. 2023. Genome-scale CRISPR-Cas9 knockout screening in nasopharyngeal carcinoma for radiosensitive and radioresistant genes. *Translational Oncology*, **30**: 101625.
Zhuang X, Qu M, Zhang X, et al. 2013. A comprehensive description and evolutionary analysis of 22 grouper (Perciformes, Epinephelidae) mitochondrial genomes with emphasis on two novel genome organizations. *PLoS One*, **8**(8): e73561.