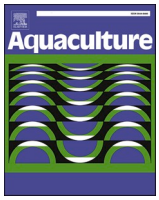
Aquaculture 613 (2026) 743301

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture



Development and evaluation of a 45 K liquid SNP array and its application in genetic improvement for spotted sea bass (*Lateolabrax maculatus*)

Chong Zhang [a,b,c], Yonghang Zhang [a,b], Shaosen Yang [d], Cong Liu [a,b], Hao Li [a,b], Yani Dong [a,b], Lingyu Wang [a,b], Xin Qi [a,b], Haishen Wen [a,b], Kaiqiang Zhang [a,b], Canming Wang [e], Yun Li [a,b,c,*]

[a] *Key Laboratory of Mariculture, Ministry of Education (KLMME), Ocean University of China, Qingdao 266003, China*
[b] *Shandong Key Laboratory of Green Mariculture and Smart Fishery, Fisheries College, Ocean University of China, Qingdao, 266003, China*
[c] *Sanya Oceanographic Institution, Ocean University of China, Sanya 572000, China*
[d] *Agro-Tech Extension Center of Guangdong Province, Guangzhou 510500, China*
[e] *Zhangzhou Youxin Fishery Co., Ltd., Zhangzhou 363000, China*

ARTICLE INFO

Keywords:
Liquid SNP array
Lateolabrax maculatus
LuXin-I
Genetic improvement

ABSTRACT

Spotted sea bass (*Lateolabrax maculatus*), a commercially vital aquaculture species in China, faces critical challenges including the degeneration of genetic characteristics and long-term generation interval which hinder sustainable industry development. Genomic selection (GS) presents a promising strategy to accelerate genetic improvement of economically crucial traits. However, the absence of cost-effective, high-throughput genotyping tools has impeded GS implementation in this species. To address this gap, in this study, the first 45 K liquid SNP array for spotted sea bass, named as "LuXin-I", was developed using genotyping by target sequencing (GBTS) technology. Leveraging high-confidence SNP dataset identified from 1107 WGS data exhibiting the highest genetic diversity, 41,604 genome-wide background SNPs with uniform genomic distribution and appropriate MAF ranges, and 3393 functional SNPs associated with economically important traits, selection signatures, and high impact genomic regions were successfully integrated in LuXin-I SNP array and used for probe design. A total of 130,563 mSNPs within 44,997 target genomic segments were captured through genotyping evaluation for 218 test samples. The SNP array demonstrated robust genotyping performance, with exceptional genotype call rates of 99.48 % and 99.52 % for core SNPs and mSNPs, and high concordance with GC and R^2 values greater than 94 %. Moreover, only 1.05 % and 0.40 % of core SNPs and mSNPs had a genotype missing rate greater than 0.1, confirming its accuracy and reliability for genotyping spotted sea bass samples. In addition, LuXin-I SNP array demonstrated identical performance in resolving population structure compared to WGS data while reduced GWAS resolution for trait-associated SNPs due to marker density limitations. Notably, the SNP array achieved comparable heritability (h^2) estimates ranging from 0.534 to 0.592 and superior genomic prediction accuracy for growth traits comparable to WGS data. In summary, LuXin-I SNP array will provide a cost-effective and reliable genotyping platform, enabling large-scale genotyping for spotted sea bass and advancing genetic improvement.

1. Introduction

Spotted sea bass (*Lateolabrax maculatus*) has emerged as a commercially vital mariculture fish in China, with annual production surpassing 200,000 tons in recent years (Li et al., 2025a). Renowned for its superior nutritional profile, delicate flavor, and cultural importance, this species has gained substantial popularity in both domestic and international markets. However, the lack of high-quality strains or varieties has consistently hindered the development and expansion of the spotted sea bass industry due to various environmental threats, disease invasions and germplasm deterioration caused by frequent inbreeding practices (Zhang et al., 2024). Genomic selection (GS) has revolutionized modern breeding programs by offering enhanced prediction accuracy, reduced inbreeding rates, accelerated genetic gains, and shortened generational intervals, which has contributed to remarkable genetic improvement for plant, livestock and aquaculture species (Crossa et al., 2017; Georges

* Corresponding author at: Key Laboratory of Mariculture, Ministry of Education (KLMME), Ocean University of China, Qingdao, 266003, China.
E-mail address: yunli0116@ouc.edu.cn (Y. Li).

https://doi.org/10.1016/j.aquaculture.2025.743301
Received 9 April 2025; Received in revised form 7 September 2025; Accepted 13 October 2025
Available online 14 October 2025

et al., 2019; Zhou et al., 2024b). Our team has committed to accelerating genetic improvement for economically important traits of spotted sea bass and has achieved several genetic advancements in the application of GS for growth (Zhang et al., 2023a), alkalinity tolerance (Zhang et al., 2025a), heat tolerance (Liu et al., 2025) and swimming performance traits (Li et al., 2025a). Through extensive GS modeling, we found that an appropriate number of informative single nucleotide polymorphism (SNP) markers is sufficient to achieve high predictive accuracy, indicating that whole-genome resequencing (WGS) generates excessive redundant markers (millions of SNPs) that provide diminishing returns for prediction accuracy. Moreover, genotyping large-scale individuals using WGS becomes economically prohibitive in commercial breeding operations. Therefore, developing high-throughput and cost-effective genotyping tools were urgent need for accelerating genetic progress of spotted sea bass through GS breeding.

While both solid SNP arrays and reduced-representation genome sequencing (RRGS) serve as cost-effective genotyping solutions for population-level studies (Zhou et al., 2020), their technical architectures present distinct trade-offs that require careful consideration in specific research methods and goals (Kockum et al., 2023). Compared to enzyme digestion bias and uneven coverage of genomic regions of RRGS methods, solid SNP arrays are favored genotyping tools for their feasibility to be customized to target specific SNPs, high genotyping accuracy, robustness and straightforward use (Liu et al., 2023; Robledo et al., 2018; Zhou et al., 2020). These merits have driven successful array development for high-quality aquaculture species including common carp (*Cyprinus carpio*) (Xu et al., 2014), Nile tilapia (*Oreochromis niloticus*) (Yáñez et al., 2020), channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*) (Liu et al., 2014; Zeng et al., 2017), Atlantic salmon (*Salmo salar*) (Houston et al., 2014), large yellow croaker (*Larimichthys crocea*) (Zhou et al., 2020) and Japanese flounder (*Paralichthys olivaceus*) (Zhou et al., 2021). Efficient and accurate high-throughput genotyping capabilities using solid arrays have revolutionized genetic analyses in aquaculture, enabling large-scale genome wide association study (GWAS) and accelerating genetic improvement through GS breeding programs (Chen et al., 2024; Zhou et al., 2024b). Despite these advancements, the high customization cost, expensive equipment and low flexibility hinder the widespread development of solid arrays for aquaculture fishes (Rasheed et al., 2017). Additionally, collected samples need to be sent to foreign sequencing facilities for genotyping using solid arrays, which also increases financial and time costs. These constraints collectively underscore the urgent need for developing affordable, locally adaptable SNP array solutions tailored to aquaculture industries.

The emergence of genotyping by target sequencing (GBTS) technology represents a novel genotyping strategy that effectively integrates the complementary advantages of solid SNP arrays and RRGS approaches. This approach firstly capture and hybridize targeted genomic regions using custom probes, followed by high-depth sequencing to accurately identify and genotype SNPs after library construction (Guo et al., 2019; Guo et al., 2021). Therefore, GBTS, also known as liquid SNP array, combines the advantages of customizability, high genotyping accuracy and repeatability of solid arrays with the flexible design and low genotyping cost of RRGS approaches (Liu et al., 2023; Wang et al., 2023). Moreover, due to reduced design costs and short turnaround times, new and non-informative genetic variants can be constantly added to or removed from liquid SNP arrays whenever necessary, conferring liquid SNP arrays with huge potential for applications in breeding research (Li et al., 2025b). Recently, liquid SNP arrays have gradually become mainstream in array research and widely applied in species such as maize (Guo et al., 2019), wheat (Xiang et al., 2023), chicken (Liu et al., 2023), pigs (Zhang et al., 2025b) and cattle (Chen et al., 2024). For aquaculture fishes, a 20 K DongXin I liquid SNP array was developed for vibriosis-resistant germplasm of the leopard coral grouer (*Plectropomus leopardus*) (Zhou et al., 2024a), a 55 K NingXin-III liquid SNP array was developed based on the 55 K NingXin-II solid SNP array for large yellow croaker (Wang et al., 2023), a 20 K liquid SNP were developed to reveals QTLs for disease resistance in tiger pufferfish (*Takifugu rubripes*) (Li et al., 2025b). In addition, several liquid SNPs were also developed for other aquaculture species, including Pacific abalone (*Haliotis discus hannai*) (Liu et al., 2022), and estuarine oyster (*Crassostrea ariakensis*) (Zhang et al., 2023b), mud crab (*Scylla paramamosain*) (Ye et al., 2025), and Pacific white shrimp (*Litopeneaus vannamei*) (Yu et al., 2020). This expanding applications of GBTS in aquatic species underscores the critical need for developing a liquid SNP array for spotted sea bass. Such a platform would not only overcome the economic constraints of traditional WGS genotype but also enable dynamic adaptation to evolving breeding objectives, ultimately accelerating genetic gain in this commercially vital species.

Leveraging GBTS technology, we developed “LuXin-I", a 45 K liquid SNP array tailored for spotted sea bass. This array integrates numerous background SNPs that are evenly distributed across the genome, as well as several functional SNPs associated with economically important traits, selection signatures, and high impact genomic regions. Furthermore, the genotyping performance of LuXin-I was first validated using 218 test samples using comprehensive criteria, and the application performance for genetic improvement was systematically compared to WGS data through population structure analysis, GWAS, genetic parameter estimations, and GP. Our study would provide an open-access, cost-effective and reliable tool for large-scale genotyping for spotted sea bass, facilitating its genetic research and GS application.

2. Materials and methods

2.1. Data description and SNP identification

To comprehensively capture the genetic diversity of spotted sea bass, we implemented a dual-strategy sampling approach. First, we selected 100 representative wild germplasm samples collected from 14 sites along the Chinese coastline (100data). These samples exhibited different genetic structures and clustered into three highly differentiated populations including Bohai Gulf (BH) and Beibu Gulf (BB) and intermediate (IM) populations. Among these, the BH and BB populations show significant phenotypic divergence in traits such as low-temperature tolerance and growth performance (Chen et al., 2023). Second, to ensure the broad capabilities of SNP arrays in genotyping diverse wild and farmed populations, we selected a total of 1007 WGS data including 301 wild and 706 farmed individuals. Specifically, 1007 samples were sourced from three local fish farms in Dongying (DY), Tangshan (TS), and Yantai (YT), China. This included 301 one-year-old fish from DY farm, collected from natural populations in the Yellow Sea and Bohai Sea (DY wild), 213 five-year-old broodstock from TS farm (TS breeding), both showing growth differentiation at the individual level. And 493 two-year-old fish from YT farm were derived from northern (fast growth strain) and southern farmed populations, which exhibiting phenotypic differentiation including growth and heat tolerance trait. Detailed sequencing information is described in **Table S1**, and variant calling were conducted using HaplotypeCaller procedure of GATK (v4.5.0.0) (McKenna et al., 2010). SNP refinement was conducted using Plink (v1.90) based on the following criteria: minor allele frequency (maf) > 0.05, genotyping call rate (geno) > 0.95, the *P*-value of the Hardy-Weinberg equilibrium test (hwe) > 0.01 (Purcell et al., 2007). Furthermore, SNPs with a heterozygosity rate above 0.5 and non-biallelic SNPs were excluded using BCFtools (v1.9) (Genovese et al., 2024) to establish a high-confidence SNP dataset for subsequent analyses and marker selection of SNP array.

2.2. Population structure and genetic diversity analysis

High-confidence SNP were subjected to population structure and genetic diversity analysis. Principal component analysis (PCA) was conducted using Plink (Purcell et al., 2007), with the first two principal components (PC1 and PC2) visualized to reveal genetic clustering. Population structure analysis was inferred using Admixture (v1.3.0)

(Alexander et al., 2009) with K values ranging from 1 to 5, allowing the software to infer population stratification based on the optimal K values. Furthermore, we calculated within-population nucleotide diversity (π) for wild, breeding and farmed population. Pairwise fixation indices (*Fst*) were computed to calculate the level of genetic differentiation between each pair of populations using VCFtools (v0.1.16) with a 10-kb sliding window and a 5-kb step size (Danecek et al., 2011).

2.3. Functional SNPs selection for LuXin-I SNP array

2.3.1. Functional SNPs identification using genome-wide association study (GWAS)

Functional SNPs underlying economically important traits are essential for the accuracy and efficiency of calculating genomic estimated breeding values (GEBV). To achieve this, we strategically prioritized several GWAS-informative SNPs associated with economic traits and incorporated them into the array development. In detail, our team has conducted GWAS using 514 samples (DY and TS individuals) for growth traits (Zhang et al., 2023a), 287 DY samples for alkalinity tolerance trait (Zhang et al., 2025a), 493 YT samples for heat tolerance trait (Liu et al., 2025), 446 samples for swimming performance trait (Li et al., 2025a), 514 samples (367 males and 147 females) for sex determination (unpublished data), and 450 samples for disease resistance to *Nocardia seriolae* (unpublished data). Detailed phenotype differentiation data is described in **Table S2** and mixed linear model (MLM) incorporating population structure and genetic relatedness was conducted to identify functional SNPs associated with these traits using GEMMA software (v0.98.5) (Zhou and Stephens, 2012). Given heterogeneous genetic architecture across traits, there were no standardized criteria for determining the number of functional SNPs to be selected. For example, for polygenic traits including growth, alkalinity tolerance, heat tolerance and swimming performance, which are regulated by numerous micro-effect SNPs rather than a few major QTLs (Li et al., 2025a; Liu et al., 2025; Zhang et al., 2023a), we initially selected approximately 500 SNPs with the highest ranked *P*-values based on the GWAS results. However, for traits like sex determination and disease resistance, where significant major QTLs were detected and most functional SNPs were within 1000-bp distance (unpublished data), we modestly reduced the number of selected SNPs to around 200. All GWAS-informative SNPs were merged, and duplicates were removed, retaining only one SNP within every 300-bp window. We set a suggestive significance threshold of *P*-value <0.0001 to select SNPs for SNP array design.

2.3.2. Functional SNPs identification using selective sweep analysis

Selection signatures are genotypic markers shaped during evolutionary adaptation under natural or artificial selection. Population structure analysis of 1107 spotted sea bass samples revealed two major genetic groups, primarily driven by differentiation between northern and southern farmed populations. To ensure the effectiveness of LuXin-I SNP array in germplasm resource management and population structure identification, we conducted selective sweep analyses, including fixation index (*Fst*) and nucleotide diversity ($\theta\pi$) using VCFtools (v0.1.13) (Danecek et al., 2011), to identify positively selected regions focused on two main comparisons. First, since the wild and breeding populations formed a single genetic cluster that overlapped with partially fast growth strain, we aimed to identify genomic regions underlying the divergence between these two major genetic groups (northern vs. southern farmed populations). Second, as the northern farmed population represents the fast growth strain, we specifically compared it against wild populations to detect artificial selection signals potentially related to growth performance. Candidate regions were defined as those within the top 5 % of *Fst* values and the extreme 2.5 % of the $\theta\pi$ ratio distribution. From these regions, all associated SNPs were merged, and only one SNP per 10 kb window was retained to minimize redundancy.

2.3.3. Functional SNPs identification based on impact effects

For categorizing the impact effects of SNPs, we firstly construct the databases for reference genome of spotted sea bass (JAYMHB000000000) using "build" mode of SnpEff software (v5.0), then we performed SNP functional annotation using "ann" mode of SnpEff according to the annotated genomic locations (Cingolani et al., 2012). Annotation results classified SNPs into 4 impact effects, including HIGH, LOW, MODERATE and MODIFIER types based on their predicted biological consequences. Of which, high-impact SNPs were defined as variants that cause significant disruptions to protein structure or function, including stop gained, frameshift, and splice site variants, highlighting their crucial roles in genomic regions. Consequently, all high-impact SNPs were prioritized for array inclusion due to their putative roles in critical biological pathways.

A composite functional SNP set was generated by integrating three types of functional SNPs, and only one SNP within each 300 bp window was retained to eliminate redundancy and ensure adequate spacing between SNPs. Finally, this comprehensive selection approach incorporated a total of 3393 functional SNPs that will enable accelerated breeding progress.

2.4. Determination of background SNPs and probe design

To ensure an even distribution of physical distance and MAF values of background SNPs across the genome, we first divided the final high-confidence SNP database into five MAF bins: 0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, and 0.4–0.5. These MAF bins were selected to capture a broad spectrum of allele frequencies, from rare to common variants, thereby ensuring comprehensive coverage of genetic diversity within diverse populations. This stratification enhances the utility of array in genomic analyses and breeding applications by providing balanced representation across different allele frequency spectra. Due to the varying number of SNPs within each MAF bin, with initial counts of 1,416,395, 1007,089, 493,439, 358,533 and 252,906 in the respective bins, we selected background SNPs within each MAF bin using different window sizes to maintain an even distribution of SNPs across the genome. Specifically, we retained only one SNP within every 35 kb, 30 kb, 20 kb, 15 kb, and 12 kb window for 0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, and 0.4–0.5 MAF bins, respectively. The selection of smaller window sizes for higher MAF bins ensures that common variants are adequately represented, while larger window sizes for lower MAF bins prevent excessive redundancy of rare variants. However, several SNPs in different MAF bins still exist in close physical distance in genome. Furthermore, we merged all background SNPs and randomly retained one SNP within each 10 kb genomic region, which could retain appropriate SNP numbers for array development. Additionally, background SNPs located within 300 bp of functional SNPs were further removed to avoid data redundancy. Our approach effectively mitigates SNP density discrepancies of different MAF bins and minimizes redundancies, ultimately resulting in a total of 41,607 background SNPs that are evenly distributed across all chromosomes.

Finally, a total of 45,000 SNPs were selected for probe design. Each probe was designed as a 110 bp double-stranded DNA sequence, adhering to quality standards of 30 %–70 % GC content for background SNPs and 20 %–80 % GC content for functional SNPs. Additionally, probes were constrained to regions with five or fewer homologous sequences to ensure specificity. Three background SNPs were removed due to probe design failures. The remaining probes were subsequently optimized and adjusted based on genotyping results. Ultimately, the "LuXin-I" liquid SNP array was developed, encompassing 44,997 target SNPs using GBTS technology for spotted sea bass.

2.5. The assessment of genotyping performance for LuXin-I SNP array

To evaluate the genotyping accuracy and reliability of LuXin-I SNP array for spotted sea bass samples with complex population contexts,

two rounds of genotyping using the 45 K SNP array were conducted on 218 representative samples, encompassing 132 north samples and 86 south samples. Genomic DNA was extracted from the pectoral fins of all individuals. After assessing DNA quantity and quality using a nucleic acid analyzer (OSTC, China), high-quality DNA samples were utilized for DNA library construction and sequencing at MolBreeding Biotechnology Co., Ltd., Shijiazhuang. The genotyping process and hard filter using GATK has been described in our previous study (Zhang et al., 2023a). To evaluate the genotyping performance of LuXin-I SNP array for core SNPs, genotype call rate, genotype missing rate were calculated using VCFtools (v0.1.13), genotype concordance (GC) and the squared Pearson correlation coefficient of genotype dosage (R^2) between the SNP array and WGS data were calculated using BCFtools (v1.20). Furthermore, to ensure consistency and reliability across the entire dataset, identical criteria were applied to evaluate the genotyping performance of multiple single-nucleotide polymorphisms (mSNPs), which captured in target genomic segments due to common features of GBTS technology (Liu et al., 2022).

2.6. The application of LuXin-I SNP array in genetic improvement for spotted sea bass

To evaluate the application potential of LuXin-I SNP array in genetic improvement of spotted sea bass, we extracted all 130,563 mSNPs from WGS data of an expanded cohort of 451 samples, which included the initial 218 representative samples. Four growth traits including body weight (BW), body height (BH), total length (TL) and body length (BL) were recorded and detailed information were provided in **Table S3**. This strategy enabled us to assess the effectiveness and applicability of both the mSNP set and the full SNP set in a larger population without incurring additional genotyping costs (Liu et al., 2022). The performance of SNP array and WGS data was compared through population structure analysis, GWAS, genetic parameter estimations, and GP. Genotype data from 451 samples were filtered using Plink software (v1.90) as following parameters: –geno 0.05, −-hwe $1e^{-5}$, and –maf 0.05, resulting in 99,968 SNPs for the SNP array and 4,760,402 SNPs for the WGS data.

To evaluate the ability of LuXin-I SNP array to accurately reflect the underlying population structure compared to WGS data, a neighbor-joining tree was constructed using genetic distances matrix calculated with VCF2Dis (v1.45) (https://github.com/BGI-shenzhen/VCF2Dis). In addition, principal component analysis (PCA) was conducted using Plink, and the first two principal components (PC1 and PC2) were used to delineate genetic groups. The GWAS analyses for four growth traits were performed based on MLM using GEMMA (v0.98.5) (Zhou and Stephens, 2012), the first two PCs and marker effect value were selected as fixed effects, and polygenic effect value was integrated in MLM as random effects to control for false positives. The results based on SNP array and WGS data were compared to assess the effectiveness of the SNP array in detecting trait-related SNPs.

To further validate the effectiveness of the SNP array in capturing genetic variance associated with economic traits. The genomic relationship matrix (GRM) for both datasets was estimated using GCTA (v1.94.0) with parameter: –make-grm (Yang et al., 2011). Heritability (h^2) for growth traits was estimated as the formula: $h^2 = \sigma g^2 / (\sigma g^2 + \sigma e^2)$, where σg^2 represents the additive genetic variance and σe^2 is the residual variance, the variance components were calculated by GCTA using parameter: –reml. In addition, genetic correlation among traits was also calculated using GCTA with parameter of "–reml-bivar". To access the application potential of LuXin-I SNP array in genomic selection, genomic prediction (GP) for BW trait was conducted using five-fold cross-validation with 5 replicates. Specifically, 451 samples were randomly divided into training sets ($n = 361$) and testing sets ($n = 50$). Of which, SVM model, one of machine-learning methods belonging to kernel-based algorithms, was used to build GS model based on training sets using the R package kernlab (v0.9–32) (Karatzoglou et al., 2004), then the GS model was used to calculate the genomic breeding values (GEBVs) of testing sets. We built 6 different SNP sets with the numbers of 0.1 k, 1 k, 5 k, 10 k and 100 k. GWAS was performed in train sets to select corresponding numbers of SNP based only on *P*-value. Furthermore, predictive accuracies were defined as the average Pearson correlation between actual phenotypes and GEBVs divided by the square root of h^2 for BW trait.

3. Results and analysis

3.1. Population structure and genetic diversity analysis

WGS data were collected from both wild, breeding and farmed populations, including 1107 individuals, to generate 3,528,641 high-quality SNPs for the development of LuXin-I SNP array. To clarify the genetic background and population stratification of 1107 individuals, we conducted population structure and genetic diversity analysis (Fig. 1). Principal component analysis (PCA) revealed that wild and breeding populations collectively formed a single genetic cluster, whereas farmed populations exhibited more complex genetic structures. Notably, southern farmed samples formed a distinct genetic group separated from other clusters along the first principal component (PC1), which explained 40.75 % of the total genetic variance (Fig. 1A). Northern farmed samples (representing fast growth strain) were further subdivided into two genetic subgroups, supported by both PCA and ADMIXTURE analysis as the optimal structure (K = 3) (Fig. 1B, **Table S4**), with several individuals still clustering within the wild and breeding populations. Furthermore, we estimated nucleotide diversity (π) within wild, breeding, and farmed populations, as well as pairwise fixation indices (*Fst*) between populations. Nucleotide diversity was relatively consistent across groups, ranging from 1.452×10^{-3} to 1.506×10^{-3}. The farmed population showed the highest π value (1.506×10^{-3}), further supporting the genetic distinction between northern and southern farmed populations. Genetic differentiation, measured by *Fst*, was low between breeding and wild populations (*Fst* = 0.00497), moderate between farmed and wild populations (*Fst* = 0.0180), and highest between farmed and breeding populations (*Fst* = 0.0212), indicating that farmed populations are more genetically differentiated from their wild and breeding counterparts (Fig. 1C).

3.2. The statistical results of the core SNPs

Building upon the high-quality SNP set that effectively captured population stratification and genetic diversity, we consider these SNPs that are widely distributed across population, thus providing a representative and unbiased genomic basis for subsequent marker selection to design for LuXin-I SNP array. The whole array design and application has been described in Fig. 2. Following an intensive filtering process and probe design, 44,997 target SNPs including 3393 functional SNPs and 41,604 background SNPs were integrated in the SNP array (Table 1). Detailed genotype data information including SNP_ID, chromosome, position, and allele have been described in **Table S5** and **Table S6** for functional and background SNPs, respectively. In addition, we have provided detailed *P*-value of GWAS-informative SNPs, categorized by their phenotype **in Supplementary Table 5**. Overall, target SNPs were evenly distributed across the genome, with an average density of one SNP per 13,838 bp, except for several high-density genomic regions on chromosome 19 related to sex determination (Fig. 3A). The density and distribution of target SNPs indicated the array's ability to broadly represent genetic diversity while targeting specific traits. Moreover, the number of MAF values was 3817, 6120, 8844, 10396 and 15,820 for the 0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, and 0.4–0.5 MAF bins, respectively (Fig. 3B). In our analysis, we retained a certain number of SNPs with low MAF values, considering that many crucial SNPs are often rare variants.

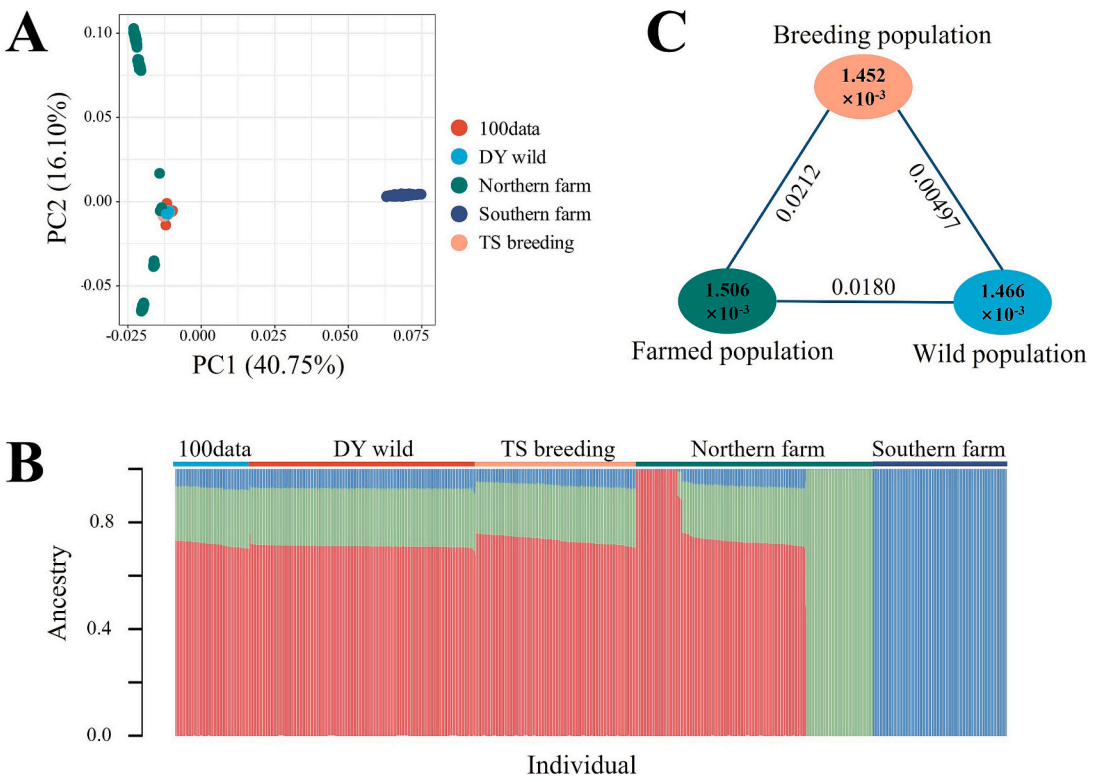


Fig. 1. (A) The principal component analysis (PCA) of all 1107 samples. (B) Population genetic structure analysis of 1107 samples based on optimal K value =3. (C) Genetic diversity analysis for wild, breeding and farmed populations, the value on each dotted line indicates fixation indices (Fst) between the two populations, and the value in each cycle represents nucleotide diversity (π) in corresponding population.

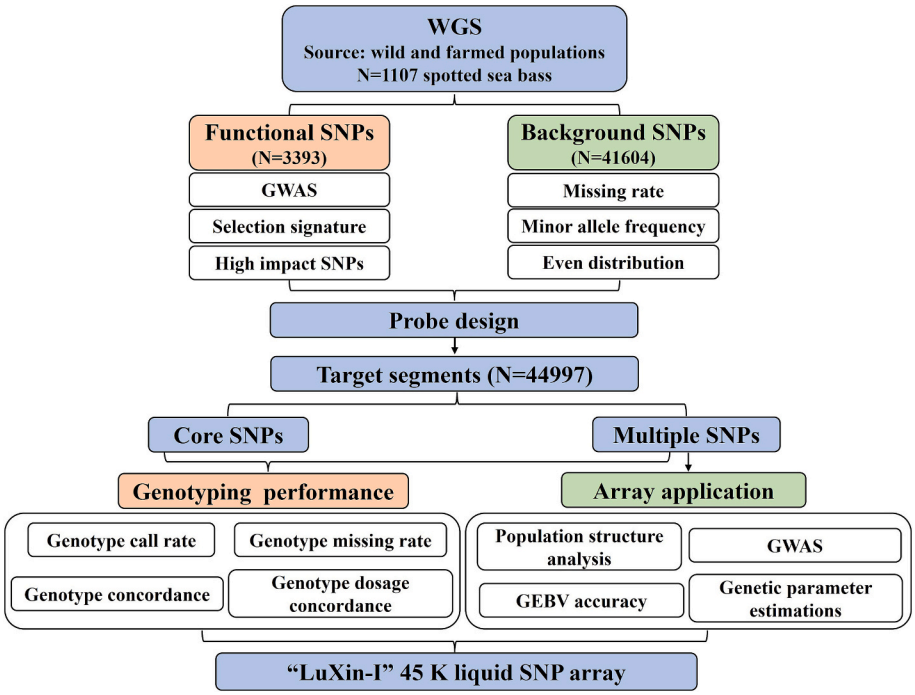


Fig. 2. Design framework of the LuXin-I SNP array and flow chart of application scenarios in genetic improvement for spotted sea bass.

3.3. Characterization and summary of the LuXin-I SNP array

Genotyping was performed on 218 test samples using LuXin-I SNP array, successfully capturing a total of 130,563 mSNPs within 44,997 target genomic segments. The distribution of mSNPs, target segments, and chromosome lengths in the SNP array was shown in Fig. 4A. Each chromosome contained approximately 1.87 K target segments and 5.44 K mSNPs, resulting in an average of 2.90 SNPs genotyped per target segment. The MAF distributions of core SNPs and mSNPs in the test samples were significantly different, with average MAF values of 0.304

Table 1
Summary of target SNPs in each type for LuXin-I SNP array.

Types of target SNP in LuXin-I SNP array	SNP number	Sample size	Data source
Background SNPs	41,604	1107	(Chen et al., 2023; Liu et al., 2025; Zhang et al., 2023a)
Growth related	441	514	(Zhang et al., 2023a)
Sex determination related	174	514	Unpublished
Nocardia seriolae resistance related	19	450	Unpublished
Heat tolerance related	445	493	(Liu et al., 2025)
Alkalinity tolerance related	444	287	(Zhang et al., 2025a)
Swimming performance related	452	446	(Li et al., 2025a)
Selective signature between new strains and wild populations	50	530	Unpublished
Selective signature between northern and southern populations	315	1107	Unpublished
High impact in genome	1053	1107	(Chen et al., 2023; Liu et al., 2025; Zhang et al., 2023a)

for core SNPs and 0.216 for mSNPs, respectively **(**Fig. 4B**, Table S7)**. This indicates that core SNPs possess higher polymorphism rates compared to mSNPs. Additionally, the SNP location of core SNPs and mSNPs in genomic region were basically identical **(**Fig. 4C**)**. Although most core SNPs and mSNPs were located in intergenic or intronic regions, a higher proportion of core SNPs (0.0564) and mSNPs (0.0460) in exon regions was observed in array data compared to WGS data (0.0269) from our previous research (Zhang et al., 2023a). These results suggest that SNP genotyping data using LuXin-I SNP array could provide adequate coverage across different genomic regions.

3.4. The evaluation of genotyping for the LuXin-I SNP array

Four metrics including genotype call and missing rate, GC and R^2 values were used to assess the genotyping performance of LuXin-I SNP array on 218 test samples. A remarkable genotype performance for core SNPs was observed, with locus calling rates ranging from 96.90 % to 99.70 %, and an average of 99.48 % (Fig. 5A**, Table S8**). The GC and R^2 values for core SNPs between array data and WGS data were 94.57 % and 94.02 %, respectively (Fig. 5A**, Table S8**). In addition, the genotype call rate for mSNPs was 99.52 %, with GC and R^2 values of 94.10 % and 94.80 %, respectively (Fig. 5B**, Table S8**). The genotype missing rates of core SNPs in the test samples indicated that 83.26 % of SNPs had no missing genotypes, and only 1.05 % of SNPs had a genotype missing rate greater than 0.1 (Fig. 5C). Although the SNP array detected a higher number of mSNPs compared to core SNPs, a lower percentage (0.40 %) was observed for mSNPs that genotype missing rates greater than 0.1 (Fig. 5D). These results indicate that most target segments effectively capture both core SNPs and mSNPs. Overall, the genotyping results demonstrate that LuXin-I SNP array is a highly accuracy and reliability genotyping tool for spotted sea bass.

3.5. Population structure analysis

In this study, 451 samples derived from northern and southern cultivated populations were selected to assess the efficacy of LuXin-I SNP array in genetic improvement for spotted sea bass. Phylogenetic trees constructed from both SNP array and WGS data indicated that the cultivated spotted sea bass samples could be significantly divided into two genetic groups: North and South (Fig. 6A). PCA results also supported the classification of samples into two subpopulations (Fig. 6B**)**. These population structure analyses demonstrate the superior performance of the LuXin-I SNP array in characterizing the population structure of spotted sea bass.

3.6. The comparison of genome-wide association results of growth traits

Given the substantial difference in variant density between LuXin-I SNP array (99,968 SNPs) and WGS data (4,760,402 SNPs), we applied platform-specific thresholds to enable a comparative analysis of detection power: 0.05/N (5.00×10^{-7}) for the SNP array and 1/N (2.10×10^{-7}) for WGS data, where N represents the total number of SNPs analyzed per platform. It is important to note that these thresholds were used for comparative purposes and do not imply strict genome-wide significance. Our objective was to evaluate the relative performance of the SNP array against WGS in detecting trait-associated signals, rather than to identify novel trait-related SNPs. For TL, BL and BW traits (Fig. 7A-C**, Table S9**), almost all trait-related SNPs identified in the WGS data were also detected by the SNP array. Although several SNPs, including SNP:1–8,192,599, SNP:1–8,192,639 and SNP:13–20,968,525, were not directly detected using SNP array, nearby SNPs (SNP:1–8,193,268 and SNP:13–20,968,579) were also identified. These proximity SNPs could also be annotated to the same genes, thereby maintaining the association with the respective traits. However, the SNP array exhibited relatively poor performance for the BH trait (Fig. 7D**, Table S9**), as several SNPs on chromosome 4 identified through WGS data were not genotyped by the SNP array. Overall, the evenly and widely spaced mSNPs in LuXin-I SNP array provide a moderate level of detection power for GWAS targeting economically important traits. However, the lower SNP density inherently limits the enough detection

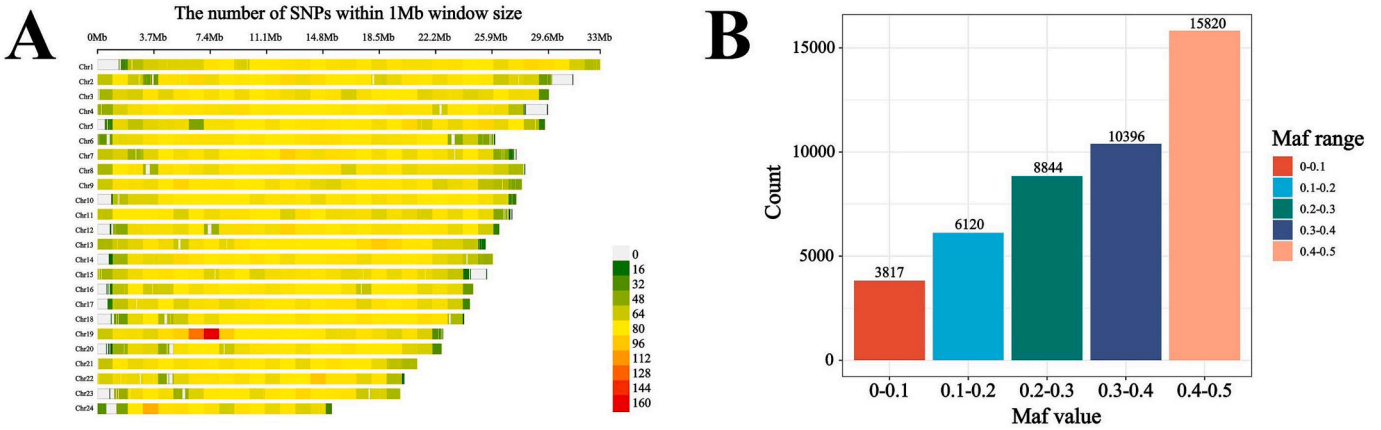


Fig. 3. (A) Genome-wide distribution of 44,997 target SNPs across 24 chromosomes. Different colors represent the corresponding number of SNPs within 1 Mb distance according to the legend. **(B)** Minor allele frequency (MAF) distribution histogram of target SNPs in the LuXin-I SNP array.

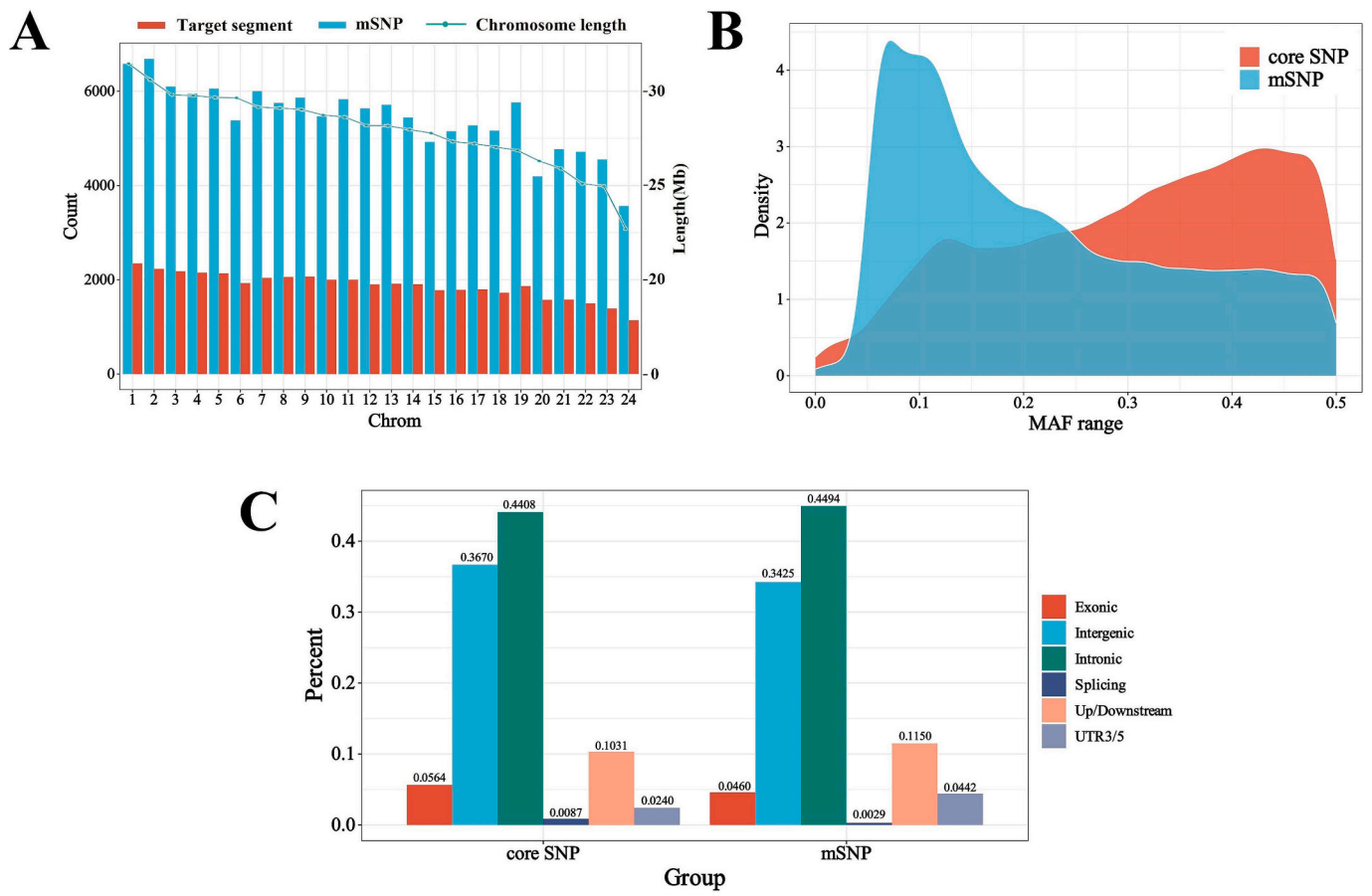


Fig. 4. **(A)** Distribution statistics of target segments and multiple single-nucleotide polymorphisms (mSNPs) in the LuXin-I SNP array across 24 chromosomes. Target segments and mSNPs counts are shown on the left axis and chromosome length is shown on the right axis. **(B)** The density distribution curve of MAF value for core SNPs and mSNPs in the LuXin-I SNP array. **(C)** Statistics of core SNPs and mSNPs with different genomic regions.

power compared to WGS data. Therefore, continuously incorporating new functional SNPs is essential to enhance the detection capabilities of LuXin-I SNP array. This ongoing enhancement will ensure that LuXin-I SNP array remains effective tools for identifying meaningful genetic associations in spotted sea bass.

3.7. Genetic parameter estimations and genomic prediction (GP)

The heritability (h^2) estimation and genetic correlations for growth traits in 451 samples were calculated using SNP array and WGS data in spotted sea bass (Table 2). Heritability estimations obtained from SNP array data were 0.581, 0.579, 0.592 and 0.550 for TL, BL, BH and BW traits, respectively. These estimates were slightly higher than those derived from WGS data, which were 0.572, 0.565, 0.571, and 0.534 for TL, BL, BH, and BW, respectively. In addition, high genetic correlations (0.694–0.992) for growth traits were observed and the results between SNP array and WGS data were basically identical. These results indicate that SNP array data are sufficient for estimating genetic variance components related to growth traits in spotted sea bass. Furthermore, predictive accuracies for BW trait at different SNP numbers were compared between SNP array and WGS data. We observed that predictive accuracies based on SNP array data (ranging from 0.686 to 0.813) were consistently higher than those based on WGS data (ranging from 0.643 to 0.681) across various SNP numbers. This trend was particularly pronounced when using 0.1 K SNPs (Fig. 8 **and Table S10)**. These findings indicated that LuXin-I SNP array is a more effective tool for selection breeding than WGS data in spotted sea bass.

4. Discussion

Reducing genotyping costs while obtaining high-quality genotype data remains a critical challenge in aquaculture breeding and genetic research. Liquid SNP array technology offers an efficient genotyping strategy and has been successfully developed for several aquaculture species. This technology enables the identification of functional genes and facilitates the selection and improvement of desirable traits such as growth rate, disease resistance, and environmental adaptability (Li et al., 2025b; Liu et al., 2022; Zhou et al., 2024b).

Spotted sea bass, a commercially important mariculture fish in China, has historically lacked high-quality varieties. To address this, we have focused on utilizing genetic resources and genomic selection to improve economically important traits in spotted sea bass over recent years. However, our previous research relied exclusively on WGS data, which poses significant genotyping cost challenges for future selection breeding programs that require genotyping large numbers of individuals. Therefore, developing a liquid SNP array is an urgent necessity to facilitate molecular design and breeding with high accuracy and efficiency. In this study, we developed a high-throughput 45 K liquid SNP array, named as “LuXin-I", based on resequencing data using GBTS technology, and evaluated its genotyping performance and application potential in the genetic improvement of spotted sea bass. To the best of our knowledge, this study represents the first development and application of a liquid SNP array for spotted sea bass.

Here, WGS data from 1107 samples were selected for SNP array development. We acknowledge the limitation regarding pedigree and parental sex information. For wild individuals, parental data was unavailable as fish samples were captured directly from natural marine

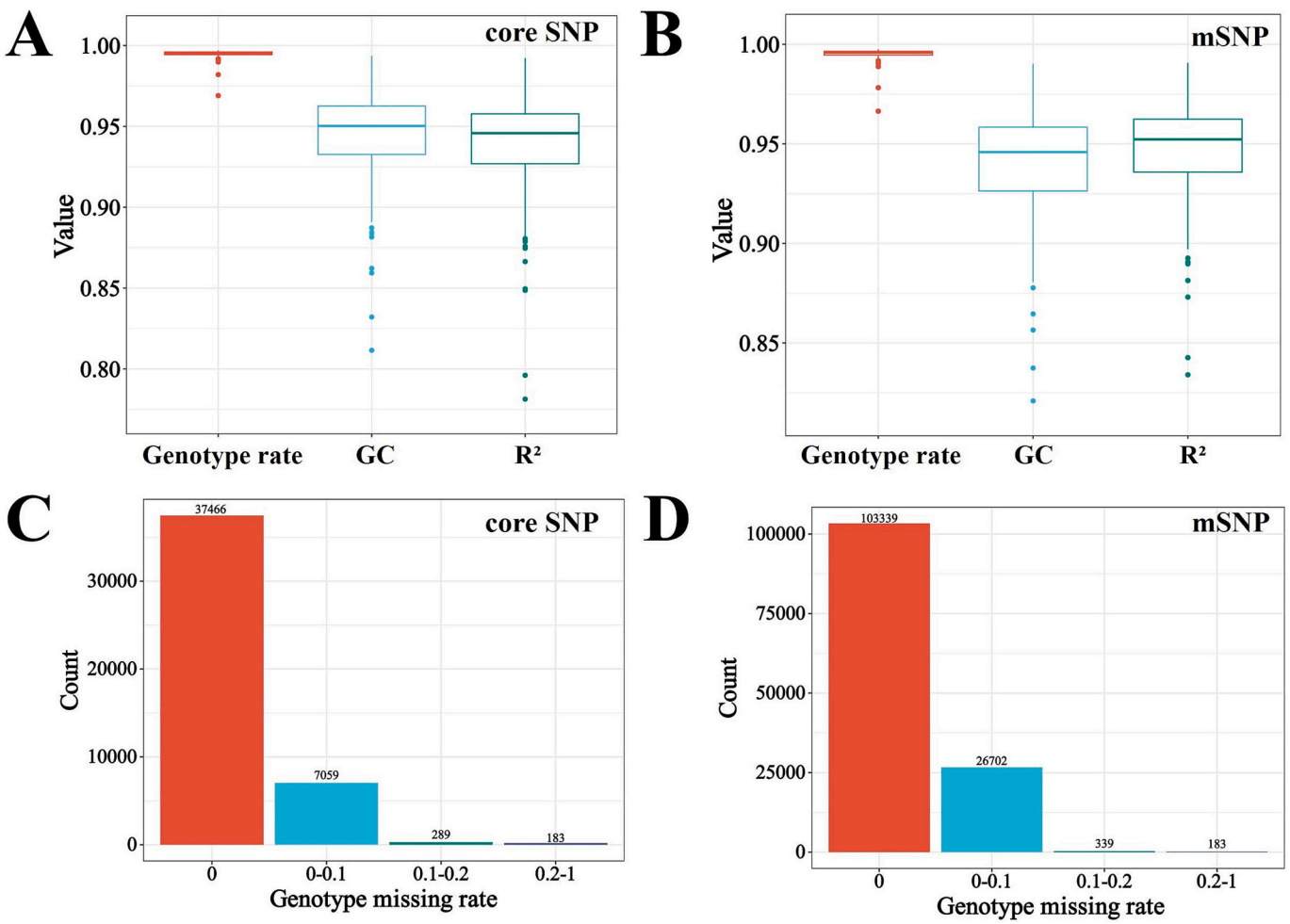


Fig. 5. Genotyping evaluation for the LuXin-I SNP array with genotype call, GC and R^2 between the SNP array and WGS data for **(A)** core SNPs and **(B)** mSNPs. GC: genotype concordance; R^2: the squared Pearson correlation coefficient of genotype dosage. Statistics of genotype missing rate for **(C)** core SNPs and **(D)** mSNPs.

environments. Practically, sexual identification of adult spotted sea bass cannot be reliably achieved through non-destructive methods during non-breeding periods, and the long breeding season (October–January) introduces breeding inconsistency for each parent fish. Furthermore, as an important mariculture fish in China, systematic genetic breeding programs for spotted sea bass remain at a nascent stage. Given the operational constraints of large-scale breeding stock management (typically >200 fish per pound) and infrastructure limitations, the commercial breeding strategy mainly employs mass selection. Consequently, we are currently unable to provide the number of female and male parents for current experiment fish and pedigree information for experimental samples. In the future, we will accurately determine parent-offspring relationships based on genotype data to advance genomic selection. In addition, population structure and genetic diversity analysis based on 3,528,641 high-quality SNP set clearly revealed comprehensive genetic background and population stratification for wild, breeding and farmed populations (Fig. 1). These further indicated that high-quality SNPs effectively capture common genetic variation present in diverse populations of spotted sea bass without introducing significant population bias, thus providing a representative and unbiased genomic basis for subsequent marker selection.

Following the principles of evenly genomic distribution and appropriate MAF ranges, we performed a rigorous filtering process to select 41,607 representative SNPs as background SNPs. Additionally, based on our research including both published and unpublished data, we incorporated 3393 functional SNPs related to economically important traits, selective signatures, and high-impact genomic regions into the array design. After excluding three background SNPs due to probe design failures, a total of 44,997 customized SNPs were selected as target SNPs for the development of LuXin-I SNP array. Genotyping results demonstrated a significant increase in the number of detectable mSNPs (130,563), with an average of 2.9 times more SNPs detected per target genomic segment (44,997), highlighting the high SNP content of LuXin-I SNP array (Fig. 4A). The distribution of MAF values revealed a higher site polymorphism for core SNPs (0.304) compared to mSNPs (0.216), as expected, given the strict filtering applied to core SNPs, while mSNPs were less stringently considered (Fig. 4B). Furthermore, the proportion of core SNPs and mSNPs across different genomic regions was nearly identical (Fig. 4C), with a higher percentage of SNPs located in exon regions compared to WGS data, further demonstrating the representativeness of SNP data genotyped by LuXin-I SNP array. High individual calling rates were observed for both core SNPs (99.48 %) and mSNPs (99.52 %), with low SNP missing rates (1.05 % and 0.40 %, respectively), highlighting the robustness of LuXin-I SNP array as a high-throughput genotyping tool (Fig. 5). These excellent genotyping performances are consistent with those of liquid SNP arrays used in aquaculture fishes such as leopard coral grouper (Zhou et al., 2024a), large yellow croaker (Wang et al., 2023) and tiger pufferfish (Li et al., 2025b). However, the GC and R^2 between SNP array and WGS data ranged from 94.10 % to 94.80 % (Fig. 5), which were relatively lower compared to those of SNP arrays for large yellow croaker (0.963) (Wang et al., 2023) and mud crab (0.954) (Ye et al., 2025). These observed concordance differences may be attributed to instances where homozygous genotypes being called as heterozygous or were there cases of opposing

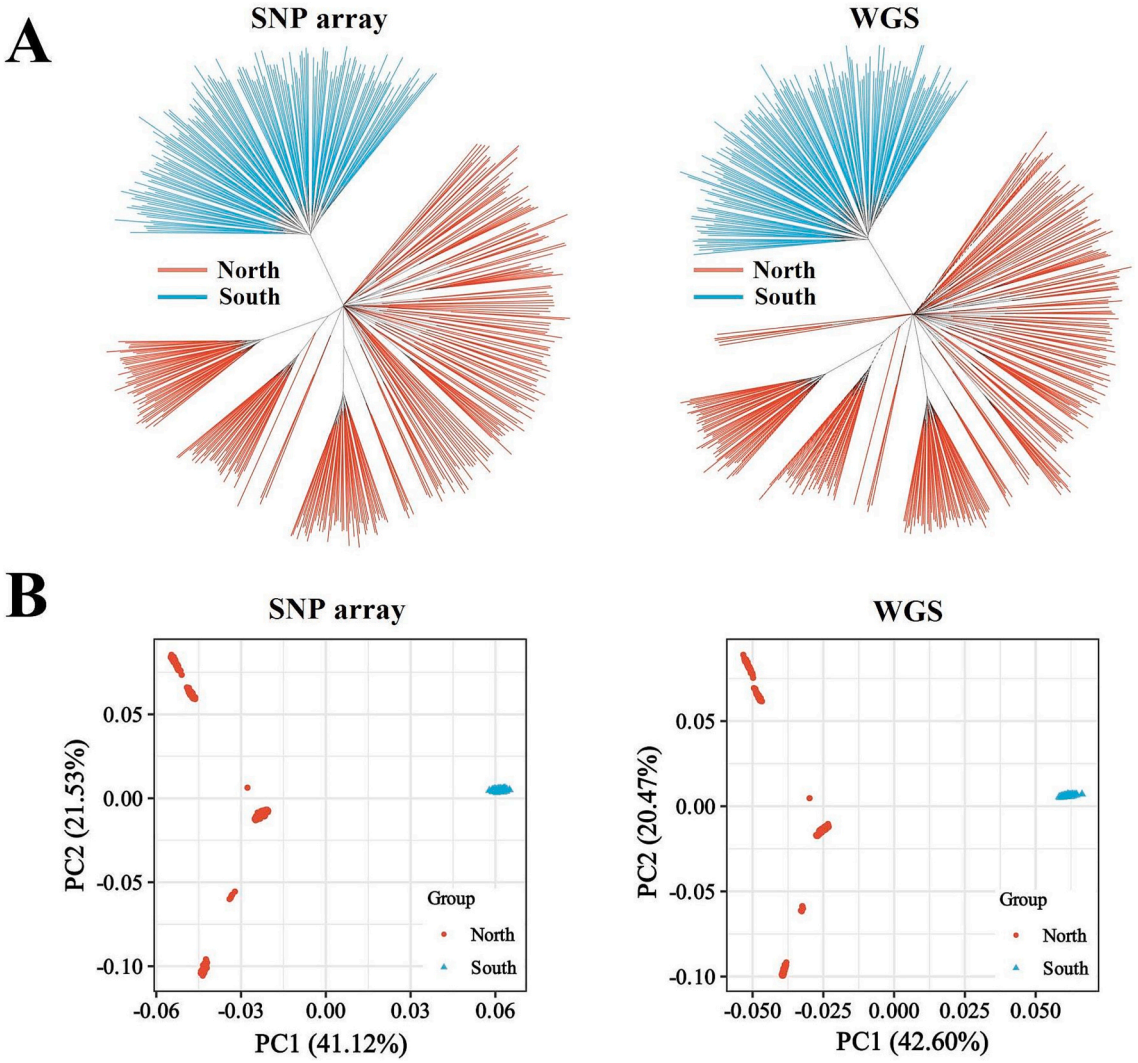


Fig. 6. The comparison of **(A)** neighbor-joining phylogenetic trees and **(B)** principal component analysis (PCA) results between SNP array and WGS data.

homozygous genotypes. We hypothesize that the uneven sequencing depth of the WGS data for several test samples, which ranged from 7.6× to 31.56× **(Table S11)**, may have contributed to genotype errors. In contrast, the SNP array employs targeted capture and high-depth sequencing of specific genomic segments, resulting in more reliable genotype calls. Despite these minor discrepancies, the overall high concordance reaffirms the accuracy and reliability of the LuXin-I SNP array for high-throughput genotyping performance.

Elucidating population structure, understanding the genetic basis of traits, and predicting GEBV for target traits are critical components of selection breeding programs for spotted sea bass. To assess the potential of LuXin-I SNP array for genetic improvement in growth traits, we systematically compared its performance across several key applications: population structure analysis, GWAS, genetic parameter estimations and GP. These comparisons offer valuable insights into the array's utility for breeding programs focused on improving the performance of economically important traits. The ability to accurately distinguish population assignments is crucial for managing genetic diversity and preventing inbreeding in aquaculture populations (Ciezarek et al., 2022). In our study, both the neighbor-joining tree and PCA, based on SNP array data, demonstrated identical performance to WGS data, effectively revealing clear genetic divergence between the northern and southern populations of spotted sea bass **(**Fig. 6**)**. However, the SNP array data exhibited lower resolution than WGS in detecting trait-associated SNPs using GWAS **(**Fig. 7**)**. Although the mSNPs genotyped by the array provide adequate genome coverage, its reduced marker density inherently constrained its ability to capture and resolve certain significant SNPs at a finer scale. Consequently, this limitation necessitates continuous updates to incorporate essential functional SNPs or the refinement of probe designs to enhance the detection capabilities of LuXin-I SNP array.

The variance components and h^2 for growth traits of spotted sea bass were first estimated using SNP information from both SNP array and WGS datasets. The high estimated h^2 values, ranging from 0.534 to 0.592 **(**Table 2**)**, confirm the substantial genetic basis of these traits, supporting the applicability of genomic selection for spotted sea bass breeding (Yu et al., 2023). Crucially, genetic correlation derived from SNP array data showed basically consistency with WGS results **(**Table 2**)**. This consistency further supports the reliability of SNP array data for quantifying genetic parameters for growth traits, making it a feasible tool for selection breeding programs. Notably, SNP array data demonstrated higher predictive accuracies for BW trait than WGS data at various marker numbers, where a minimal panel of 100 GWAS-informative SNPs achieved the maximum prediction accuracy of 0.602 for BW trait of spotted sea bass **(**Fig. 8**)**, which is significantly higher than that using WGS data (Zhang et al., 2023a). Genomic prediction accuracy serves as a crucial evaluation metric in GS, directly determining the efficiency of genetic improvement programs (Shan et al., 2021). The enhanced predictive performance observed in this study compared to previous investigations can be primarily attributed to more test samples, high genetic relatedness, and diverse population structure

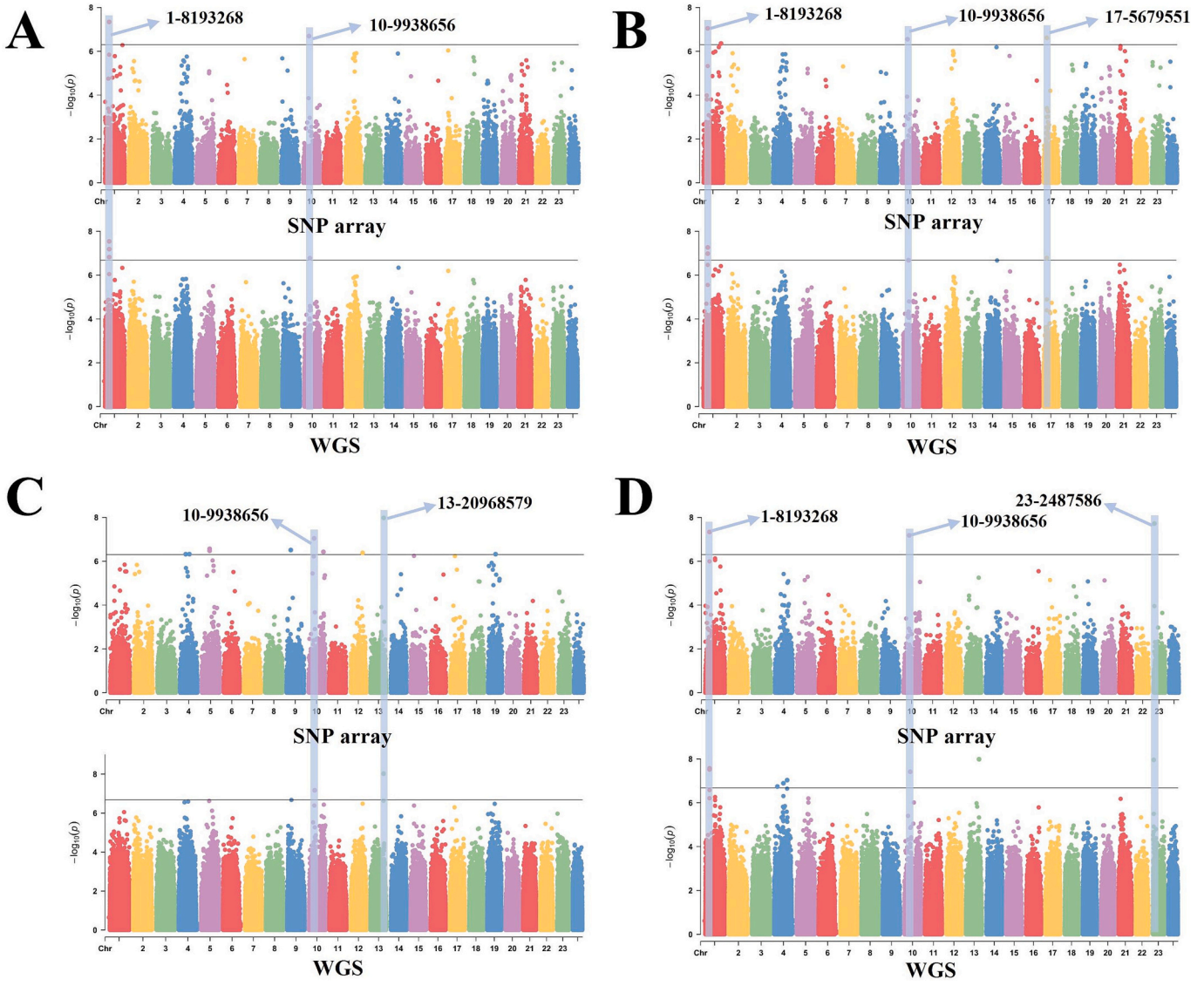


Fig. 7. The comparison of GWAS results for TL **(A)**, BL **(B)**, BW **(C)**, and BH **(D)** traits between SNP array and WGS data. Mixed linear model (MLM) was used for these analyses.

Table 2
Heritability (h^2) and genetic correlation for growth traits using genomic relationship matrix (GRM), and phenotypic correlation for growth trait using Pearson's chi-squared test. Heritability is on the diagonal in bold; Genetic and phenotypic correlations are below the diagonal and above, respectively.

Group	h^2 ± SE	TL	BL	BH	BW
SNP array data	**TL**	**0.581 ± 0.0819**	0.992 + 0.00431	0.705 + 0.0684	0.797 + 0.0592
	BL	0.981	**0.579 ± 0.0840**	0.694 + 0.0695	0.836 + 0.0537
	BH	0.793	0.797	**0.592 ± 0.0800**	0.915 + 0.0388
	BW	0.830	0.833	0.857	**0.550 ± 0.0869**
WGS data	**TL**	**0.572 ± 0.0853**	0.989 + 0.00353	0.712 + 0.0732	0.789 + 0.0602
	BL	0.981	**0.565 ± 0.0877**	0.701 + 0.0703	0.825 + 0.0556
	BH	0.793	0.797	**0.571 ± 0.0836**	0.912 + 0.0268
	BW	0.830	0.833	0.857	**0.534 ± 0.0906**

as described in (Liu et al., 2025), which is sufficient to estimate SNP effects and accurately predict GEBV of validation individuals. These findings underscore the critical importance of optimized reference population design to improve GS efficacy for growth traits in future spotted sea bass breeding program. Notably, the higher heritability estimates, and predictive accuracies obtained with the SNP array compared to the WGS dataset are reasonable and likely reflect the advantage of using a smaller set of representative SNP markers. This effect may be attributed to sufficient genetic information though relatively lower marker density, which naturally reduced model complexity and lower multicollinearity among markers relative to the extensive WGS SNP set, particularly beneficial for limited sample sizes. These situations were also observed in the comparison between InDel and SNP-based GP, where InDel markers have equivalent power to SNP markers in genetic analyses, while obtained a higher predictive accuracy due to lower marker density than SNP (Zhang et al., 2024). Overall, the better performance in genetic parameter estimations and genomic prediction of LuXin-I SNP array confirm that it could be used as a cost-effective tool for genomic selection in breeding programs of spotted sea bass.

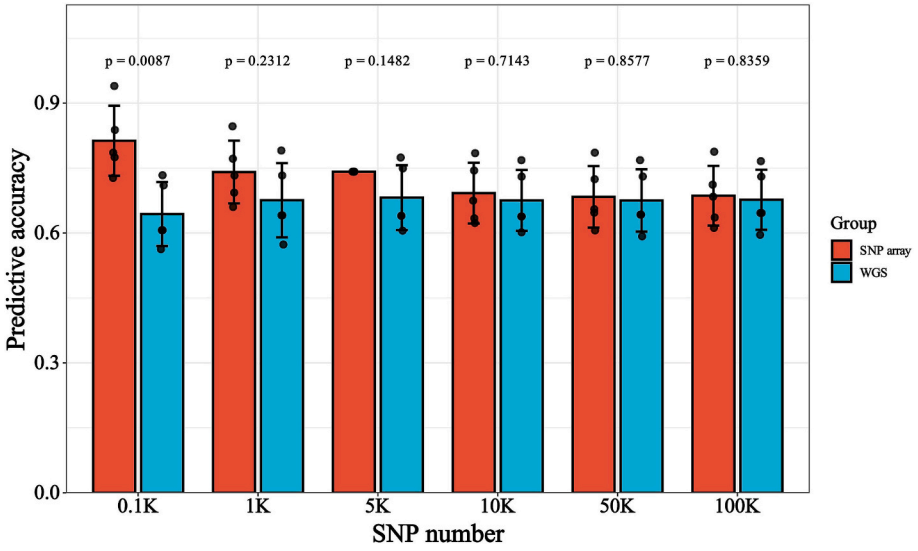


Fig. 8. The comparison of predictive accuracies for BW trait between SNP array and WGS data.The SVM model was used for genomic prediction.

5. Conclusion

In this study, the first 45 K liquid SNP array for spotted sea bass, named as "LuXin-I", was developed based on 1107 WGS data using GBTS technology. LuXin-I SNP array integrates 41,604 genome-wide background SNPs with uniform genomic distribution and appropriate MAF ranges, and 3393 functional SNPs associated with economically important traits, selection signatures, and high impact genomic regions. In addition, LuXin-I SNP array demonstrates excellent genotyping performance and application potential in population structure analysis, GWAS, genetic parameter estimations, and GP, achieving higher accuracy to WGS data. As an open-access, cost-effective and reliable genotyping platform, LuXin-I array enables large-scale genotyping for spotted sea bass and precise estimation of GEBV in genomic selection, significantly advancing genetic improvement for spotted sea bass.

CRediT authorship contribution statement

Chong Zhang: Writing – original draft, Software, Methodology, Conceptualization. **Yonghang Zhang:** Visualization, Software. **Shaosen Yang:** Resources, Funding acquisition, Conceptualization. **Cong Liu:** Visualization, Software. **Hao Li:** Visualization, Software. **Yani Dong:** Visualization, Software. **Lingyu Wang:** Visualization, Software. **Xin Qi:** Resources, Conceptualization. **Haishen Wen:** Resources, Funding acquisition, Conceptualization. **Kaiqiang Zhang:** Resources, Methodology. **Canming Wang:** Resources, Methodology. **Yun Li:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the support of the High-Performance Biological Supercomputing Center at the Ocean University of China for this research.

This research was funded by the National Key Research and Development Program of China [grant number: 2022YFD2400503], Key Research and Development Program of Shandong Province, China [grant number: 2024LZGC005], the Fundamental Research Funds for the Central Universities [grant number: 202461043], China Agriculture Research System (CARS for Marine Fish Culture Industry) [grant number: CARS-47], and research on breeding technology of candidate species for Guangdong modern marine ranching [grant number: 2024-MRB-00-001].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aquaculture.2025.743301.

Data availability

Data will be made available on request.

References

Alexander, D., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19 (9), 1655–1664. https://doi.org/10.1101/gr.094052.109.

Chen, B., Zhou, Z., Shi, Y., Gong, J., Li, C., Zhou, T., Li, Y., Zhang, D., Xu, P., 2023. Genome-wide evolutionary signatures of climate adaptation in spotted sea bass inhabiting different latitudinal regions. Evol. Appl. 16, 1029–1043. https://doi.org/10.1111/eva.13551.

Chen, Y., Guo, Y., Ge, F., Gao, H., Zhou, J., Wu, X., Qian, C., Wang, Z., Wang, Z., Zhu, B., Xu, L., Gao, X., Zhang, L., Gao, H., Li, J., 2024. Developing a liquid capture chip to accelerate the genetic progress of cattle. Anim. Res. One Health 2, 204–216. https://doi.org/10.1002/aro2.58.

Ciezarek, A., Ford, A.G.P., Etherington, G.J., Kasozi, N., Malinsky, M., Mehta, T.K., Penso-Dolfin, L., Ngatunga, B.P., Shechonge, A., Tamatamah, R., Haerty, W., Di Palma, F., Genner, M.J., Turner, G.F., 2022. Whole genome resequencing data enables a targeted SNP panel for conservation and aquaculture of *Oreochromis cichlid* fishes. Aquaculture 548, 737637. https://doi.org/10.1016/j.aquaculture.2021.737637.

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Luan, W., Land, S., Lu, X., Ruden, D., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff. Fly. 6, 80–92. https://doi.org/10.4161/fly.19695.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K., 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22, 961–975. https://doi.org/10.1016/j.tplants.2017.08.011.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Genomes Project Analysis, G, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Genovese, G., Rockweiler, N.B., Gorman, B.R., Bigdeli, T.B., Pato, M.T., Pato, C.N., Ichihara, K., McCarroll, S.A., 2024. BCFtools/liftover: an accurate and

comprehensive tool to convert genetic variants across genome assemblies. Bioinformatics 40, btae038. https://doi.org/10.1093/bioinformatics/btae038.
Georges, M., Charlier, C., Hayes, B., 2019. Harnessing genomic information for livestock improvement. Nat. Rev. Genet. 20, 135–156. https://doi.org/10.1038/s41576-018-0082-2.
Guo, Z., Wang, H., Tao, J., Ren, Y., Xu, C., Wu, K., Zou, C., Zhang, J., Xu, Y., 2019. Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. Mol. Breed. 39, 37. https://doi.org/10.1007/s11032-019-0940-4.
Guo, Z., Yang, Q., Huang, F., Zheng, H., Sang, Z., Xu, Y., Zhang, C., Wu, K., Tao, J., Prasanna, B.M., Olsen, M.S., Wang, Y., Zhang, J., Xu, Y., 2021. Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. Plant Commun. 2, 100230. https://doi.org/10.1016/j.xplc.2021.100230.
Houston, R.D., Taggart, J.B., Cézard, T., Bekaert, M., Lowe, N.R., Downing, A., Talbot, R., Bishop, S.C., Archibald, A.L., Bron, J.E., Penman, D.J., Davassi, A., Brew, F., Tinch, A.E., Gharbi, K., Hamilton, A., 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics 15, 90. https://doi.org/10.1186/1471-2164-15-90.
Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab - an S4 package for kernel methods in R. J. Stat. Softw. 11, 1–20.
Kockum, I., Huang, J., Stridh, P., 2023. Overview of genotyping technologies and methods. Curr. Protocol. 3, e727. https://doi.org/10.1002/cpz1.727.
Li, H., Zhang, C., Wen, H., Qi, X., Dong, Y., Liu, C., Zhang, Y., Niu, C., Li, Y., 2025a. Genetic basis analysis and genome prediction of swimming performance traits in juvenile spotted sea bass (*Lateolabrax maculatus*). Aquaculture 598, 741962. https://doi.org/10.1016/j.aquaculture.2024.741962.
Li, Y., Fraslin, C., Chi, Y., Mukiibi, R., Hu, Y., Wang, J., Li, W., Li, W., Houston, R.D., Robledo, D., Chen, S., 2025b. A newly developed 20 K SNP array reveals QTLs for disease resistance to Cryptocaryon irritans in tiger pufferfish (*Takifugu rubripes*). Aquaculture 595, 741652. https://doi.org/10.1016/j.aquaculture.2024.741652.
Liu, S., Sun, L., Li, Y., Sun, F., Jiang, Y., Zhang, Y., Zhang, J., Feng, J., Kaltenboeck, L., Kucuktas, H., Liu, Z., 2014. Development of the catfish 250K SNP array for genome-wide association studies. BMC. Res. Notes 7, 135. https://doi.org/10.1186/1756-0500-7-135.
Liu, J., Peng, W., Yu, F., Lin, W., Shen, Y., Yu, W., Gong, S., Huang, H., You, W., Luo, X., Ke, C., 2022. Development and validation of a 40-K multiple-SNP array for Pacific abalone (*Haliotis discus hannai*). Aquaculture 558, 738393. https://doi.org/10.1016/j.aquaculture.2022.738393.
Liu, Y., Shan, Y., Tu, Y., Zhang, M., Ji, G., Ju, X., Shi, S., Fan, C., Li, Y., Shu, J., 2023. Designing and evaluating a cost-effective single nucleotide polymorphism liquid array for Chinese native chickens. Anim. Res. One Health 1, 168–179. https://doi.org/10.1002/aro2.31.
Liu, C., Wen, H., Zhang, C., Zhang, Y., Wang, L., Sun, D., Liu, M., Dong, Y., Li, P., Li, H., Zhang, K., Qi, X., Li, Y., 2025. Deciphering the genetic basis and genomic prediction of heat tolerance trait from whole-genome resequencing in spotted sea bass (*Lateolabrax maculatus*). Aquaculture 598, 741951. https://doi.org/10.1016/j.aquaculture.2024.741951.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M., 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.
Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R.K., He, Z., 2017. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Mol. Plant 10, 1047–1064. https://doi.org/10.1016/j.molp.2017.06.008.
Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2018. Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev. Aquac. 10, 670–682. https://doi.org/10.1111/raq.12193.
Shan, X., Xu, T., Ma, Z., Zhang, X., Ruan, Z., Chen, J., Shi, Q., You, X., 2021. Genome-wide association improves genomic selection for ammonia tolerance in the orange-spotted grouper (*Epinephelus coioides*). Aquaculture 533, 736214. https://doi.org/10.1016/j.aquaculture.2020.736214.
Wang, J., Miao, L., Chen, B., Zhao, J., Ke, Q., Pu, F., Zhou, T., Xu, P., 2023. Development and evaluation of liquid SNP array for large yellow croaker (*Larimichthys crocea*). Aquaculture 563, 739021. https://doi.org/10.1016/j.aquaculture.2022.739021.
Xiang, M., Liu, S., Wang, X., Zhang, M., Yan, W., Wu, J., Wang, Q., Li, C., Zheng, W., He, Y., Ge, Y., Wang, C., Kang, Z., Han, D., Zeng, Q., 2023. Development of breeder chip for gene detection and molecular-assisted selection by target sequencing in wheat. Mol. Breed. 43, 13. https://doi.org/10.1007/s11032-023-01359-3.
Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y., Kuang, Y., Zhang, Y., Feng, J., Li, C., Yu, J., Li, Q., Zhu, Y., Liu, Y., Xu, P., Sun, X., 2014. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). BMC Genomics 15, 307. https://doi.org/10.1186/1471-2164-15-307.
Yáñez, J.M., Yoshida, G., Barria, A., Palma-Véjares, R., Travisany, D., Díaz, D., Cáceres, G., Cádiz, M.I., López, M.E., Lhorente, J.P., Jedlicki, A., Soto, J., Salas, D., Maass, A., 2020. High-throughput single nucleotide polymorphism (SNP) discovery and validation through whole-genome resequencing in Nile Tilapia (*Oreochromis niloticus*). Mar. Biotechnol. 22, 109–117. https://doi.org/10.1007/s10126-019-09935-5.
Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
Ye, S., Zhou, X., Ouyang, M., Cui, W., Xiang, Z., Zhang, Y., Yuan, Y., Ikhwanuddin, M., Li, S., Zheng, H., Zhang, Y., Ma, H., 2025. Development and validation of a 40 K liquid SNP array for the mud crab (*Scylla paramamosain*). Aquaculture 594, 741394. https://doi.org/10.1016/j.aquaculture.2024.741394.
Yu, Y., Luo, Z., Wang, Q., Zhang, Q., Zhang, X., Xiang, J., Li, F., 2020. Development of high throughput SNP genotyping approach using target sequencing in Pacific white shrimp and its application for genetic study. Aquaculture 528, 735549. https://doi.org/10.1016/j.aquaculture.2020.735549.
Yu, H., Sui, M., Yang, Z., Cui, C., Hou, X., Liu, Z., Wang, X., Dong, X., Zhao, A., Wang, Y., Huang, X., Hu, J., Bao, Z., 2023. Deciphering the genetic basis and prediction genomic estimated breeding values of heat tolerance in Zhikong scallop *Chlamys farreri*. Aquaculture 565, 739090. https://doi.org/10.1016/j.aquaculture.2022.739090.
Zeng, Q., Fu, Q., Li, Y., Waldbieser, G., Bosworth, B., Liu, S., Yang, Y., Bao, L., Yuan, Z., Li, N., Liu, Z., 2017. Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence. Sci. Rep. 7, 40347. https://doi.org/10.1038/srep40347.
Zhang, C., Wen, H., Zhang, Y., Zhang, K., Qi, X., Li, Y., 2023a. First genome-wide association study and genomic prediction for growth traits in spotted sea bass (*Lateolabrax maculatus*) using whole-genome resequencing. Aquaculture 566, 739194. https://doi.org/10.1016/j.aquaculture.2022.739194.
Zhang, K., Li, A., Qi, H., Yang, Q., Du, M., Wang, X., Zhang, Z., Wang, C., Wang, W., Zhang, G., Li, L., 2023b. The development of a 30 K SNP genotyping tool targeting genomic regions of temperature and salinity adaptation in estuarine oyster. Aquaculture 566, 739168. https://doi.org/10.1016/j.aquaculture.2022.739168.
Zhang, Z, Xing, S, Qiu, A, Zhang, N, Wang, W, Qian, C, Zhang, J, Wang, C, Zhang, Q, Ding, X., 2025b. The development of a porcine 50K SNP panel using genotyping by target sequencing and its application. J Inergr Arg 24 (5), 1930–1943. https://doi.org/10.1016/j.jia.2023.07.033.
Zhang, C., Zhang, Y., Liu, C., Wang, L., Dong, Y., Sun, D., Wen, H., Zhang, K., Qi, X., Li, Y., 2024. Genome-wide association study and genomic prediction for growth traits in spotted sea bass (*Lateolabrax maculatus*) using insertion and deletion markers. Anim. Res. One Health 2 (4), 400–416. https://doi.org/10.1002/aro2.87.
Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–824. https://doi.org/10.1038/ng.2310.
Zhang, Y., Zhang, C., Wen, H., Qi, X., Wang, Q., Zhang, K., Wang, L., Sun, D., Dong, Y., Li, P., Li, Y., 2025a. Genetic basis and identification of candidate genes for alkalinity tolerance trait in spotted sea bass (*Lateolabrax maculatus*) by genome-wide association study (GWAS). Mar. Biotechnol. 27, 27. https://doi.org/10.1007/s10126-024-10405-w.
Zhou, T., Chen, B., Ke, Q., Zhao, J., Pu, F., Wu, Y., Chen, L., Zhou, Z., Bai, Y., Pan, Y., Gong, J., Zheng, W., Xu, P., 2020. Development and evaluation of a high-throughput single-nucleotide polymorphism array for large yellow croaker (*Larimichthys crocea*). Front. Genet. 11. https://doi.org/10.3389/fgene.2020.571751.
Zhou, Q., Chen, Y.-D., Lu, S., Liu, Y., Xu, W.-T., Li, Y.-Z., Wang, L., Wang, N., Yang, Y.-M., Chen, S.-L., 2021. Development of a 50K SNP array for Japanese flounder and its application in genomic selection for disease resistance. Engineering 7, 406–411. https://doi.org/10.1016/j.eng.2020.06.017.
Zhou, Q., Lu, S., Liu, Y., Zhou, B., Chen, S., 2024a. Development of a 20 K SNP array for the leopard coral grouper, *Plectropomus leopardus*. Aquaculture 578, 740079. https://doi.org/10.1016/j.aquaculture.2023.740079.
Zhou, Q., Wang, J., Li, J., Chen, Z., Wang, N., Li, M., Wang, L., Si, Y., Lu, S., Cui, Z., Liu, X., Chen, S., 2024b. Decoding the fish genome opens a new era in important trait research and molecular breeding in China. Sci. China Life Sci. 67, 2064–2083. https://doi.org/10.1007/s11427-023-2670-5.